PATENT APPLICATION TRANSMITTAL LETTER
(Large Entity)

Docket No.
1038-921 MIS:jb

TO THE ASSISTANT COMMISSIONER FOR PATENTS

Transmitted herewith for filing under 35 U.S.C. 111 and 37 C.F.R. 1.53 is the patent application of:

Sheena M. Loosmore; Ken Sasaki; Yan Ping Yang; and Michel H. Klein

For: **RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA**

Enclosed are:

- ☐ Certificate of Mailing with Express Mail Mailing Label No.
- ☒ **Sixty-Eight (68)** sheets of drawings.
- ☐ A certified copy of a application.
- ☒ Declaration ☐ Signed. ☒ Unsigned.
- ☐ Power of Attorney
- ☐ Information Disclosure Statement
- ☐ Preliminary Amendment
- ☐ Other:

CLAIMS AS FILED

For	#Filed	#Allowed	#Extra		Rate	Fee
Total Claims	23	- 20 =	3	x	$18.00	$54.00
Indep. Claims	8	- 3 =	5	x	$78.00	$390.00
Multiple Dependent Claims (check if applicable) ☐						$0.00
					BASIC FEE	$760.00
					TOTAL FILING FEE	$1,204.00

☒ A check in the amount of **$1,204.00** to cover the filing fee is enclosed.

☒ The Commissioner is hereby authorized to charge and credit Deposit Account No. as described below. A duplicate copy of this sheet is enclosed.

- ☐ Charge the amount of as filing fee.
- ☒ Credit any overpayment.
- ☒ Charge any additional filing fees required under 37 C.F.R. 1.16 and 1.17.
- ☐ Charge the issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance, pursuant to 37 C.F.R. 1.311(b).

Dated: **July 26, 1999**

Michael Stewart
Michael I. Stewart *Signature*
(24,973)

cc:

TITLE OF INVENTION

RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA

FIELD OF INVENTION

The present invention relates to the field of immunology and is particularly concerned with outer membrane proteins from *Moraxella*, methods of recombinant production thereof, genes encoding such proteins and uses thereof.

BACKGROUND OF THE INVENTION

Otitis media is the most common illness of early childhood with approximately 70% of all children suffering at least one bout of otitis media before the age of seven. Chronic otitis media can lead to hearing, speech and cognitive impairment in children. It is caused by bacterial infection with *Streptococcus pneumoniae* (approximately 50%), non-typable *Haemophilus influenzae* (approximately 30%) and *Moraxella (Branhamella) catarrhalis* (approximately 20%). In the United States alone, treatment of otitis media costs between one and two billion dollars per year for antibiotics and surgical procedures, such as tonsillectomies, adenoidectomies and insertion of tympanostomy tubes. Because otitis media occurs at a time in life when language skills are developing at a rapid pace, developmental disabilities specifically related to learning and auditory perception have been documented in youngsters with frequent otitis media.

M. catarrhalis mainly colonizes the respiratory tract and is predominantly a mucosal pathogen. Studies using cultures of middle ear fluid obtained by tympanocentesis have shown that *M. catarrhalis* causes approximately 20% of cases of otitis media (ref. 1 - Throughout this application, various references are referred to in parenthesis to more fully describe the

state of the art to which this invention pertains. Full bibliographic information for each citation is found at the end of the specification, immediately preceding the claims. The disclosures of these references are hereby incorporated by reference into the present disclosure).

The incidence of otitis media caused by *M. catarrhalis* is increasing. As ways of preventing otitis media caused by pneumococcus and non-typable *H. influenzae* are developed, the relative importance of *M. catarrhalis* as a cause of otitis media can be expected to further increase.

M. catarrhalis is also an important cause of lower respiratory tract infections in adults, particularly in the setting of chronic bronchitis and emphysema (refs. 2, 3, 4, 5, 6, 7, and 8). *M. catarrhalis* also causes sinusitis in children and adults (refs. 9, 10. 11, 12, and 13) and occasionally causes invasive disease (refs. 14, 15, 16, 17, 18, and 19).

Like other Gram-negative bacteria, the outer membrane of *M. catarrhalis* consists of phospholipids, lipopolysaccharide (LPS), and outer membrane proteins (OMPs). Eight of the *M. catarrhalis* OMPs have been identified as major components. These are designated by letters A to H, beginning with OMP A which has a molecular mass of 98 kDa to OMP H which has a molecular mass of 21 kDa (ref. 20).

Recently, Klingman and Murphy purified and characterized a high molecular-weight outer membrane protein of *M. catarrhalis* (ref. 21). The apparent molecular mass of this protein varies from 350 kDa to 720 kDa as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). This protein appears to be an oligomer of much smaller proteins or subunits thereof of molecular mass about 120 to 140 kDa and is antigenically conserved among strains of *Moraxella*.

Helminen et al also identified a protein of molecular mass of about 300 to 400 kDa, named UspA, that was reported to be present on the surface of *Moraxella* (ref. 22).

In WO 96/34960 and US Patent No. 5,808,024, assigned to the assignee hereof and the disclosure of which is incorporated herein by reference, there is described a new protein of *M. catarrhalis* which had an apparent molecular mass of about 200 kDa. Western blot analysis using antiserum raised against the 200 kDa protein suggested that this protein was different from the large UspA protein (> 300 kDa), reported by the two groups in refs. 21 and 22. Recently, the gene sequences encoding two related proteins, UspA1 and UspA2, have been published (ref. 23). A sequence comparison between the two genes encoding the UspA proteins and the gene encoding the 200 kDa protein confirmed that the 200 kDa protein is different from either of the UspA1 and UspA2 proteins.

Fitzgerald et al (ref. 29) have identified a 200 kDa protein associated with haemagglutination. Transmission electron microcopy studies (ref. 30) showed that the 200 kDa protein associated with haemagglutination is present on the outer fibrillar layer of *M. catarrhalis*. Recently, a non-clumping variant of strain 4223 was prepared by serial passaging and it was observed that the non-clumping variant had reduced expression of both UspA and a 200 kDa protein that is not UspA (ref. 31). It is possible that this 200 kDa protein is the same as that described in WO 96/34960 and herein.

The 200 kDa protein described herein has been detected in most, but not all, strains of *Moraxella catarrhalis*, which have been isolated from various sources, including otitis media (OM), sputum, nasopharynx, expectorate and bronchial secretions. Table 1A below contains a listing of *M. catarrhalis* strains

tested, their source and whether or not the 200 kDa protein is expressed.

M. catarrhalis infection may lead to serious disease. It would be advantageous to provide recombinant means for providing large quantities of 200 kDa outer membrane protein of *M. catarrhalis* strains and genes encoding such proteins from various *M. catarrhalis* strains for use as antigens in immunogenic preparations including vaccines, carriers for other antigens and immunogens and the generation of diagnostic reagents.

SUMMARY OF THE INVENTION

The present invention is directed towards the provision of a recombinantly-produced purified and isolated outer membrane protein of *Moraxella catarrhalis* and other *Moraxella* strains, having an apparent molecular mass of about 200 kDa, as well as genes encoding the same from various strains of *Moraxella catarrhalis*.

In one aspect of the present invention, there is provided an isolated and purified nucleic acid molecule having (a) a nucleotide sequence set forth in Figure 3, 4 or 5 (SEQ ID Nos: 5, 6, 8, 9, 11, 12) for *Moraxella catarrhalis* strains 4223, Q8 and LES-1 respectively or the complementary sequence thereto; (b) a nucleotide sequence encoding an about 200 kDa outer membrane protein of a strain of *Moraxella catarrhalis* and having the derived amino acid sequence shown in Figures 3, 4 or 5 (SEQ ID Nos: 7, 10, 13) for *Moraxella catarrhalis* strains 4223, Q8 and LES-1 respectively; and (c) a nucleotide sequence encoding an about 200 kDa outer membrane protein of another strain of *Moraxella catarrhalis* which is characterized by a tract of consecutive G nucleotides which is 3 or a multiple thereof in length, an ATG start codon about 80 to 90 bp upstream of said tract and said tract being located between about amino acids 25 and 35 encoded by the nucleotide sequence.

The another strain of *Moraxella catarrhalis* in (c) is a strain as identified in Table 1A other than strains 4223, Q8 and LES-1 and expressing an about 200 kDa protein.

In another aspect of the invention, there is provided (a) a nucleotide sequence set forth in Figure 8 (SEQ ID No: 12) for a 5'-truncation of the gene encoding an about 200 kDa outer membrane protein of *Moraxella catarrhalis* strain 4223; (b) a nucleotide sequence encoding the derived amino acid sequence set forth in Figure 9 (SEQ ID No: 13) for a N-terminal truncation of an about 200 kDa outer membrane protein of *Moraxella catarrhalis* strain 4223; and (c) a nucleotide sequence encoding a 5'-truncation of a gene encoding an about 200 kDa outer membrane protein of another strain of *Moraxella catarrhalis* and being capable of expressing the corresponding N-terminally truncated about 200 kDa outer membrane protein from *E. coli*.

A further aspect of the invention providing an isolated and purified nucleic acid molecule which is a contiguous *Nde* I - *Pst* I fragment of SEQ ID No: 5.

The invention, in an additional aspect, provides a vector for transforming a host comprising a nucleic acid molecule as provided herein, which may be a plasmid vector. The plasmid vector may be one which has the identifying characteristics of pKS348 (ATCC 203,529) or pKS294 (ATCC 203,528). The plasmid vector also may be one having the identifying characteristics of pQWE or pQWF.

A further aspect of the invention provides a host cell, such as *E. coli*, transformed by a vector provided herein and expressing an about 200 kDa protein of a strain of *Moraxella catarrhalis* or an approximately C-terminal half thereof. The invention further provides, in an additional aspect, a recombinant about 200 kDa outer membrane protein of a strain of *Moraxella*

catarrhalis or an approximately C-terminal half thereof producible by the transformed host provided herein.

The recombinant about 200 kDa outer membrane protein or an approximately C-terminal half thereof may be formulated into an immunogenic composition, which may be formulated as a vaccine for *in vivo* administration to protect against disease caused by *Moraxella catarrhalis*, which may be provided in combination with a targeting molecule for delivery to specific cells of the immune system, formulated as a microparticle, capsule or liposome preparation, and may further comprise an adjuvant.

The invention, in a further aspect, includes a method of inducing protection against disease caused by *Moraxella catarrhalis* by administering to a susceptible host, which may be a human, an effective amount of the immunogenic composition provided herein.

In an additional aspect, the invention provides a method for the production of an about 200 kDa outer membrane protein of a strain *of Moraxella catarrhalis* or an approximately C-terminal half thereof, which comprises:

transforming a host cell, such as *E. coli*, with a vector as provided herein,

growing the host cell to express the encoded about 200 kDa protein or an approximately C-terminal half thereof, and

isolating and purifying the expressed about 200 kDa protein or an approximately C-terminal half thereof.

The encoded about 200 kDa protein may be expressed in inclusion bodies. The isolation and purification of the about 200 kDa protein may be effected by:

disrupting the grown transformed cells to produce supernatant and the inclusion bodies,

solubilizing the inclusion bodies to produce a solution of the recombinant about 200 kDa protein,

chromotographically purifying the solution of recombinant about 200 kDa protein free from contaminating proteins, and

isolating the purified recombinant about 200 kDa protein.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 shows restriction maps of subclones of a gene encoding the 200 kDa outer membrane protein of *M. catarrhalis* from λEMBL3 clone 8II and the location of PCR primers used to amplify the 5'-region of the gene. The open reading frame of the about 200 kDa outer membrane protein is indicated by the shaded box. The numbers in parenthesis are approximate sizes of DNA inserts in plasmids. Restrictions sites are Sal: *Sal*I, N: *Nco*I, B: *Bgl*II, K: *Kpn*I, Xb: *Xba*I, Xh: *Xho*I, RV: *Eco*RV;

Figure 2 shows the nucleotide sequence (SEQ ID No: 1 - entire sequence, SEQ ID No: 2 - coding sequence) of the gene encoding the about 200 kDa outer membrane protein of *M. catarrhalis* strain 4223, as determined from λEMBL3 clone 8II, and deduced amino acid sequence (SEQ ID No: 3 - identified GTG start codon, SEQ ID No: 4 - putative ATG start codon shaded) of the about 200 kDa outer membrane protein. A ten-G nucleotide segment of the 5'-UTR is identified by underlining. An ATG start codon for the same sequence but with a nine-G nucleotide segment is identified by a shaded box (see Figure 3);

Figure 3 shows the nucleotide sequence (SEQ ID No: 5 - entire sequence, SEQ ID No: 6 - coding sequence) of the gene encoding the about 200 kDa outer membrane protein of *M. catarrhalis* strain 4223, as determined from PCR-amplified genomic DNA of strain 4223 and the deduced amino acid sequence (SEQ ID No: 7) of the corresponding about 200 kDa outer membrane protein. A nine-G nucleotide segment of the sequence corresponding to the 10-G nucleotide segment of Figure 2, is

identified by underlining. The GTG start codon identified in Figure 2 is identified by a light box;

Figure 4 shows the nucleotide sequence (SEQ ID No: 8) of the gene encoding the about 200 kDa outer membrane protein of *M. catarrhalis* strain Q8 and the deduced amino acid sequence (SEQ ID No: 9) of the corresponding about 200 kDa outer membrane protein. A nine-G nucleotide segment is identified by underlining;

Figure 5 shows the nucleotide sequence (SEQ ID No: 10) of the gene encoding the about 200 kDa outer membrane protein of *M. catarrhalis* strain LES-I and the deduced amino acid sequence (SEQ ID No: 11) of the corresponding about 200 kDa outer membrane protein. A three-G nucleotide segment is identified by underlining;

Figure 6 contains an alignment of the amino acid sequence (in single letter code) of the about 200 kDa proteins of *M. catarrhalis* strain 4223 (SEQ ID No: 7), Q8 (SEQ ID No: 9) and LES-I (SEQ ID No: 11). The alignments of the sequences were made using BLAST and manual methods and are compared to the 4223 sequence. Gaps in the sequence where no corresponding or related amino acid exists are designated by "-" while identical amino acids are designed by ".";

Figure 7 shows the restriction sites of the *M. catarrhalis* strain 4223 derived 200 kDa protein gene as well as the identity of various plasmids containing partial or full length 200 kDa genes;

Figure 8 shows the nucleotide sequence (SEQ ID No: 12) and deduced amino acid sequence (SEQ ID No: 13) of the 5'-truncated gene encoding the M56 200 kDa protein of *M. catarrhalis* strain 4223 contained in pKS348;

Figures 9A and 9B contain a schematic of the procedure for producing plasmid pKS294 expressing the full length 200 kDa protein of *M. catarrhalis* strain 4223;

Figure 10 is a schematic of the procedure for producing plasmid pKS348 expressing the N-truncated M56 r200 kDa protein of *M. catarrhalis* strain 4223;

Figure 11 shows a schematic procedure for the purification of recombinantly-produced 200 kDa protein from *E. coli*;

Figure 12 shows SDS-PAGE analysis of the expression of M56 r200 kDa protein gene from *E. coli*. *M. catarrhalis* strain 4223 lysate was run as a positive control (a) and uninduced KS358 cultured overnight was run as a negative control (b). In each lane, 20 μg of total protein was loaded;

Figure 13 shows the SDS-PAGE analysis of the purification of the M56 r200 kDa protein according to the scheme of Figure 11. Lane 1, *E. coli* whole cells; Lane 2, soluble proteins after 50 mM Tris/NaCl, pH8, extraction; Lane 3, soluble proteins after Tris/Triton X-100/EDTA extraction; Lane 4, soluble proteins after Tris/OG extraction; Lane 5, pellet after Tris/OG extraction; Lanes 6, 7, purified 200 kDa protein;

Figure 14 shows the anti-M56 r200 kDa protein antibody titers obtained in mice. Mice were immunized on day 1, day 29 and day 43 with 0.3 μg, 1 μg, 3 μg or 10 μg of the purified M56 r200 kDa protein in adjuvant. Antisera were obtained on days 14, 28, 42 and 56 and anti-M56 r200 kDa protein IgG titers were determined. The reactive titers of antisera were defined as the reciprocal of the dilution consistently showing a two-fold increasing in absorbance over that obtained with the pre-bleed serum sample collected on day 0;

Figure 15 shows the anti-M56 r200 kDa antibody titers in guinea pigs. Guinea pigs were immunized and antisera were analyzed according to the protocol of Figure 14;

Figure 16 shows the location of PCR primers used to amplify a DNA fragments carrying portions of the 200 kDa protein gene from chromosomal DNA of *M. catarrhalis*

strain RH408, a spontaneous mutant of strain 4223 which does not produce the 200 kDa protein;

Figure 17 is a partial nucleotide and derived amino acid sequence for the 200 kDa protein of *M. catarrhalis* strain 4223, indicating by arrows the locations of the initial amino acid of the respective three truncations ALA^{12}, VAL^{19} and GLY^{39};

Figure 18 shows schematic diagrams for two 3' half clones of the 4223 200 kDa gene. Clone pQWE contains a fusion between the 5' end of the 200 kDa gene and the 3' half of the gene. Clone pQWE contains the 3' half of the gene alone. The location of the PCR primers used to generate pQWF is indicated.

Figure 19 is a construction diagram for producing plasmid pQWE expressing a C-terminal portion of the 200 kDa protein of *M. catarrhalis* strain 4223 fused to the N-terminus; and

Figure 20 is a construction diagram for producing plasmid pQWF expressing a C-terminal portion of the 200 kDa protein of *M. catarrhalis* strain 4223.

GENERAL DESCRIPTION OF THE INVENTION

In WO 96/34960 (Figure 6), the sequence of a cloned gene from *M. catarrhalis* 4223 encoding an about 200 kDa protein, was described. The open reading frame was predicted to start at a GTG codon. Sequence analysis of 200 kDa genes from additional strains, suggested that a slightly longer open reading frame was more generally found. A re-examination of the sequence from the lambda phage-derived 200 kDa gene confirmed the GTG start codon and an upstream stretch of 10 G nucleotides in a G tract. However, when sequence analysis was performed on 4223 genomic PCR-amplified subclones, the longer open reading frame was found starting from an ATG codon. The G-tract was found to contain 9 G nucleotides in the chromosomal gene. An additional G nucleotide had been inserted during cloning from the phage library. Analysis of the 5' end of the 200 kDa gene from 24 strains

suggests that the number of G nucleotides in the G tract acts as regulator of expression.

Utilizing the techniques described herein, the genes encoding the about 200 kDa protein from *M. catarrhalis* strains Q8 and LES-1 have been cloned and sequenced. Figures 4 and 5 show respectively the nucleotide and derived amino acid sequences. An amino acid sequence comparison of the derived amino acid sequences of the 200 kDa protein from the three strains of *M. catarrhalis* is contained in Figure 6.

Based on the sequence information, a plasmid (pKS294) was constructed that contained the full-length 200 kDa protein gene of strain 4223 starting at the ATG codon, under control of the bacteriophage T7 promoter. However, even a basal level of expression of the full-length gene from the ATG was lethal to *E. coli*. Deletion of a 165 bp 5' fragment of the 200 kDa coding region greatly reduced the toxicity of the resultant protein to *E. coli*. Plasmid pKS348 contains the T7 promoter transcriptionally driving a 200 kDa protein gene which starts at amino acid residue 56. The V56 codon was changed to M56. The M56 r200 kDa protein was produced and the purified protein was used to generate guinea pig antiserum.

In WO 96/34960, a bactericidal antibody assay was described that was used to demonstrate that anti-200 kDa antibody was bactericidal for *M. catarrhalis*. The assay was used herein to demonstrate broad bactericidal antibody activity against heterologous clinical isolates from different geographical locations, by anti-M56 r200 kDa antibody. A single anti-M56 r200 kDa antibody was lytic for 62% of strains tested.

The 200 kDa protein was originally identified as a putative adhesin when its presence was detected in a clumping strain, but not a non-clumping derivative. In order to determine whether it were truly an adhesin, an *in vitro* adherence assay was developed in which the

inhibition of binding by antibody between *M. catarrhalis* and epithelial cells was measured. Using this assay, anti-M56 r200 kDa antibody was capable of inhibiting adherence of the homologous strain by 48%, demonstrating that the 200 kDa protein was an adhesin. When an additional 25 strains of *M. catarrhalis* were assayed, 21 were found to have reduced adherence to epithelial cells in the presence of anti-M56 r200 kDa antibody. 19 of these strains had not been killed by the same antibody. Thus, a single anti-M56 r200 kDa antibody was capable of killing or blocking adherence of 91% of the strains tested.

The sequence comparison for the 200 kDa gene from three strains of *M. catarrhalis* showed that the C-terminal half of the protein was quite conserved. Strain LES-1 contained an insert of about 300 amino acids. Thus, based upon the C-terminal region, the strains may be divided into two families depending upon whether they contained the insert 4223 and Q8 formed one family while LES-1 formed the other. The carboxy terminal halves (3' halves) of the 4223 or LES-1 200 kDa genes were expressed in *E. coli* with good yields and the purified carboxy terminal half of the proteins were used to generate antibodies. When tested in the bactericidal antibody assay, these antisera were bactericidal, as seen in Table 1B.

It is clearly apparent to one skilled in the art, that the various embodiments of the present invention have many applications in the fields of vaccination, diagnosis, treatment of *Moraxella* infections, and in the generation of immunological reagents. A further non-limiting discussion of such uses is further presented below.

1. Vaccine Preparation and Use

Immunogenic compositions, including those suitable to be used as vaccines, may be prepared from the about 200 kDa outer membrane protein as disclosed herein, as

well as immunological fragments and fusions thereof, which may be purified from the bacteria or which may be produced recombinantly. The vaccine elicits an immune response in a subject which produces antibodies, including anti-200 kDa outer membrane protein antibodies and antibodies that are opsonizing or bactericidal. Should the vaccinated subject be challenged by *Moraxella* or other bacteria that produce proteins capable of producing antibodies that specifically recognize 200 kDa outer membrane protein, the antibodies bind to and inactivate the bacterium. Furthermore, opsonizing or bactericidal anti-200 kDa outer membrane protein antibodies may also provide protection by alternative mechanisms.

Immunogenic compositions including vaccines may be prepared as injectables, as liquid solutions or emulsions. The about 200 kDa outer membrane protein may be mixed with pharmaceutically acceptable excipients which are compatible with the about 200 kDa outer membrane protein. Such excipients may include, water, saline, dextrose, glycerol, ethanol, and combinations thereof. The immunogenic compositions and vaccines may further contain auxiliary substances, such as wetting or emulsifying agents, pH buffering agents, or adjuvants to enhance the effectiveness thereof. Immunogenic compositions and vaccines may be administered parenterally, by injection subcutaneously or intramuscularly. Alternatively, the immunogenic compositions formed according to the present invention, may be formulated and delivered in a manner to evoke an immune response at mucosal surfaces. Thus, the immunogenic composition may be administered to mucosal surfaces by, for example, the nasal or oral (intragastric) routes. Alternatively, other modes of administration including suppositories and oral formulations may be desirable. For suppositories, binders and carriers may include, for example,

polyalkalene glycols or triglycerides. Oral formulations may include normally employed incipients such as, for example, pharmaceutical grades of saccharine, cellulose and magnesium carbonate. These compositions can take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain about 1 to 95% of the about 200 kDa outer membrane protein. The immunogenic preparations and vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective, protective and immunogenic. The quantity to be administered depends on the subject to be treated, including, for example, the capacity of the individual's immune system to synthesize antibodies, and if needed, to produce a cell-mediated immune response. Precise amounts of active ingredient required to be administered depend on the judgement of the practitioner. However, suitable dosage ranges are readily determinable by one skilled in the art and may be of the order of micrograms of the about 200 kDa outer membrane protein. Suitable regimes for initial administration and booster doses are also variable, but may include an initial administration followed by subsequent administrations. The dosage may also depend on the route of administration and will vary according to the size of the host.

The immunogenic preparations including vaccines may comprise as the immunostimulating material a nucleotide vector comprising at least a portion of the gene encoding the about 200 kDa protein, or the at least a portion of the gene may be used directly for immunization.

The concentration of the about 200 kDa outer membrane antigen in an immunogenic composition according to the invention is in general about 1 to 95%. A vaccine which contains antigenic material of only one pathogen is a monovalent vaccine. Vaccines which

contain antigenic material of several pathogens are combined vaccines and also belong to the present invention. Such combined vaccines contain, for example, material from various pathogens or from various strains of the same pathogen, or from combinations of various pathogens.

Immunogenicity can be significantly improved if the antigens are co-administered with adjuvants, commonly used as 0.05 to 0.1 percent solution in phosphate-buffered saline. Adjuvants enhance the immunogenicity of an antigen but are not necessarily immunogenic themselves. Adjuvants may act by retaining the antigen locally near the site of administration to produce a depot effect facilitating a slow, sustained release of antigen to cells of the immune system. Adjuvants can also attract cells of the immune system to an antigen depot and stimulate such cells to elicit immune responses.

Immunostimulatory agents or adjuvants have been used for many years to improve the host immune responses to, for example, vaccines. Intrinsic adjuvants, such as lipopolysaccharides, normally are the components of the killed or attenuated bacteria used as vaccines. Extrinsic adjuvants are immunomodulators which are typically non-covalently linked to antigens and are formulated to enhance the host immune responses. Thus, adjuvants have been identified that enhance the immune response to antigens delivered parenterally. Some of these adjuvants are toxic, however, and can cause undesirable side-effects, making them unsuitable for use in humans and many animals. Indeed, only aluminum hydroxide and aluminum phosphate (collectively commonly referred to as alum) are routinely used as adjuvants in human and veterinary vaccines. The efficacy of alum in increasing antibody responses to diphtheria and tetanus toxoids is well established and a HBsAg vaccine has been adjuvanted with alum. While the usefulness of alum is

well established for some applications, it has limitations. For example, alum is ineffective for influenza vaccination and inconsistently elicits a cell mediated immune response.

A wide range of extrinsic adjuvants can provoke potent immune responses to antigens. These include saponins complexed to membrane protein antigens (immune stimulating complexes), pluronic polymers with mineral oil, killed mycobacteria in mineral oil, Freund's complete adjuvant, bacterial products, such as muramyl dipeptide (MDP) and lipopolysaccharide (LPS), as well as lipid A, and liposomes.

To efficiently induce humoral immune responses (HIR) and cell-mediated immunity (CMI), immunogens are typically emulsified in adjuvants. Many adjuvants are toxic, inducing granulomas, acute and chronic inflammations (Freund's complete adjuvant) FCA, cytolysis (saponins and Pluronic polymers) and pyrogenicity, arthritis and anterior uveitis (LPS and MDP). Although FCA is an excellent adjuvant and widely used in research, it is not licensed for use in human or veterinary vaccines because of its toxicity.

Desirable characteristics of ideal adjuvants include:

(1) lack of toxicity;

(2) ability to stimulate a long-lasting immune response;

(3) simplicity of manufacture and stability in long-term storage;

(4) ability to elicit both CMI and HIR to antigens administered by various routes, if required;

(5) synergy with other adjuvants;

(6) capability of selectively interacting with populations of antigen presenting cells (APC);

(7) ability to specifically elicit appropriate T_H1 or T_H2 cell-specific immune responses; and

(8) ability to selectively increase appropriate antibody isotype levels (for example, IgA) against antigens.

U.S. Patent No. 4,855,283 granted to Lockhoff et al on August 8, 1989 which is incorporated herein by reference thereto, teaches glycolipid analogues including N-glycosylamides, N-glycosylureas and N-glycosylcarbamates, each of which is substituted in the sugar residue by an amino acid, as immuno-modulators or adjuvants. Thus, Lockhoff et al. (US Patent No. 4,855,283 and ref. 27) reported that N-glycolipid analogs displaying structural similarities to the naturally-occurring glycolipids, such as glycosphospholipids and glycoglycerolipids, are capable of eliciting strong immune responses in both herpes simplex virus vaccine and pseudorabies virus vaccine. Some glycolipids have been synthesized from long chain-alkylamines and fatty acids that are linked directly with the sugars through the anomeric carbon atom, to mimic the functions of the naturally occurring lipid residues.

U.S. Patent No. 4,258,029 granted to Moloney, assigned to the assignee hereof and incorporated herein by reference thereto, teaches that octadecyl tyrosine hydrochloride (OTH) functioned as an adjuvant when complexed with tetanus toxoid and formalin inactivated type I, II and III poliomyelitis virus vaccine. Also, Nixon-George et al. (ref. 24), reported that octadecyl esters of aromatic amino acids complexed with a recombinant hepatitis B surface antigen, enhanced the host immune responses against hepatitis B virus.

Lipidation of synthetic peptides has also been used to increase their immunogenicity. Thus, Wiesmuller (ref. 25) describes a peptide with a sequence homologous to a foot-and-mouth disease viral protein coupled to an adjuvant tripalmityl-S-glyceryl-cysteinylserylserine, being a synthetic analogue of the N-terminal part of the

lipoprotein from Gram negative bacteria. Furthermore, Deres et al. (ref. 26) reported *in vivo* priming of virus-specific cytotoxic T lymphocytes with synthetic lipopeptide vaccine which comprised of modified synthetic peptides derived from influenza virus nucleoprotein by linkage to a lipopeptide, N-palmityl-S-2,3-bis(palmitylxy)-(2RS)-propyl-[R]-cysteine (TPC).

2. Immunoassays

The about 200 kDa outer membrane protein of the present invention is useful as an immunogen for the generation of anti-200 kDa outer membrane protein antibodies, as an antigen in immunoassays including enzyme-linked immunosorbent assays (ELISA), RIAs and other non-enzyme linked antibody binding assays or procedures known in the art for the detection of anti-bacterial, anti-*Moraxella*, and anti-200 kDa outer membrane protein antibodies. In ELISA assays, the about 200 kDa outer membrane protein is immobilized onto a selected surface, for example, a surface capable of binding proteins such as the wells of a polystyrene microtiter plate. After washing to remove incompletely adsorbed about 200 kDa outer membrane protein, a nonspecific protein such as a solution of bovine serum albumin (BSA) that is known to be antigenically neutral with regard to the test sample may be bound to the selected surface. This allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific bindings of antisera onto the surface.

The immobilizing surface is then contacted with a sample, such as clinical or biological materials, to be tested in a manner conducive to immune complex (antigen/antibody) formation. This may include diluting the sample with diluents, such as solutions of BSA, bovine gamma globulin (BGG) and/or phosphate buffered saline (PBS)/Tween. The sample is then allowed to incubate for from 2 to 4 hours, at temperatures such as

of the order of about 20° to 37°C. Following incubation, the sample-contacted surface is washed to remove non-immunocomplexed material. The washing procedure may include washing with a solution, such as PBS/Tween or a borate buffer. Following formation of specific immunocomplexes between the test sample and the bound about 200 kDa outer membrane protein, and subsequent washing, the occurrence, and even amount, of immunocomplex formation may be determined by subjecting the immunocomplex to a second antibody having specificity for the first antibody. If the test sample is of human origin, the second antibody is an antibody having specificity for human immunoglobulins and in general IgG. To provide detecting means, the second antibody may have an associated activity such as an enzymatic activity that will generate, for example, a colour development upon incubating with an appropriate chromogenic substrate. Quantification may then be achieved by measuring the degree of colour generation using, for example, a visible spectrophotometer.

3. Use of Sequences as Hybridization Probes

The nucleotide sequences of the present invention, comprising the sequence of the about 200 kDa protein gene, now allow for the identification and cloning of the about 200 kDa protein gene from any species of *Moraxella*.

The nucleotide sequences comprising the sequence of the about 200 kDa protein gene of the present invention are useful for their ability to selectively form duplex molecules with complementary stretches of other about 200 kDa protein genes. Depending on the application, a variety of hybridization conditions may be employed to achieve varying degrees of selectivity of the probe toward the other genes. For a high degree of selectivity, relatively stringent conditions are used to form the duplexes, such as low salt and/or high temperature conditions, such as provided by 0.02 M to

0.15 M NaCl at temperatures of between about 50°C to 70°C. For some applications, less stringent hybridization conditions are required such as 0.15 M to 0.9 M salt, at temperatures ranging from between about 20°C to 55°C. Hybridization conditions can also be rendered more stringent by the addition of increasing amounts of formamide, to destabilize the hybrid duplex. Thus, particular hybridization conditions can be readily manipulated, and will generally be a method of choice depending on the desired results. In general, convenient hybridization temperatures in the presence of 50% formamide are: 42°C for a probe which is 95 to 100% homologous to the target fragment, 37°C for 90 to 95% homology and 32°C for 85 to 90% homology.

In a clinical diagnostic embodiment, the nucleic acid sequences of the about 200 kDa protein genes of the present invention may be used in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including radioactive, enzymatic or other ligands, such as avidin/biotin and digoxigenin-labelling, which are capable of providing a detectable signal. In some diagnostic embodiments, an enzyme tag such as urease, alkaline phosphatase or peroxidase, instead of a radioactive tag may be used. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with samples containing about 200 kDa protein gene sequences.

The nucleic acid sequences of the about 200 kDa protein genes of the present invention are useful as hybridization probes in solution hybridizations and in embodiments employing solid-phase procedures. In embodiments involving solid-phase procedures, the test

DNA (or RNA) from samples, such as clinical samples, including exudates, body fluids (e. g., serum, amniotic fluid, middle ear effusion, sputum, bronchoalveolar lavage fluid) or even tissues, is adsorbed or otherwise affixed to a selected matrix or surface. The fixed, single-stranded nucleic acid is then subjected to specific hybridization with selected probes comprising the nucleic acid sequences of the about 200 kDa protein encoding genes or fragments or analogs thereof of the present invention under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required depending on, for example, the G+C contents, type of target nucleic acid, source of nucleic acid, size of hybridization probe etc. Following washing of the hybridization surface so as to remove non-specifically bound probe molecules, specific hybridization is detected, or even quantified, by means of the label. It is preferred to select nucleic acid sequence portions which are conserved among species of *Moraxella*. The selected probe may be at least 18bp and may be in the range of about 30 to 90 bp.

4. Expression of the about 200 kDa Protein Gene

Plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell may be used for the expression of the genes encoding the about 200 kDa protein in expression systems. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* may be transformed using pBR322 which contains genes for ampicillin and tetracycline resistance and thus provides an easy means for identifying transformed cells. The plasmids or phage, must also contain, or be modified to contain, promoters which can be used by the host cell for expression of its own proteins.

In addition, phage vectors containing replicon and control sequences that are compatible with the host can be used as a transforming vector in connection with these hosts. For example, the phage in lambda GEM™-11 may be utilized in making recombinant phage vectors which can be used to transform host cells, such as *E. coli* LE392.

Promoters commonly used in recombinant DNA construction include the β-lactamase (penicillinase) and lactose promoter systems and other microbial promoters, such as the T7 promoter system as described in U.S. Patent No. 4,952,496. Details concerning the nucleotide sequences of promoters are known, enabling a skilled worker to ligate them functionally with genes. The particular promoter used will generally be a matter of choice depending upon the desired results. Hosts that are appropriate for expression of the about 200 kDa protein genes, fragments, analogs or variants thereof, may include *E. coli, Bacillus* species, *Haemophilus*, fungi, yeast, *Bordetella*, or the baculovirus expression system may be used.

In accordance with this invention, it is preferred to make the protein by recombinant methods, particularly when the naturally occurring about 200 kDa protein as purified from a culture of a species of *Moraxella* may include trace amounts of toxic materials or other contaminants. This problem can be avoided by using recombinantly produced protein in heterologous systems which can be isolated from the host in a manner to minimize contaminants in the purified material. Particularly desirable hosts for expression in this regard include Gram positive bacteria which do not have LPS and are, therefore, endotoxin free. Such hosts include species of *Bacillus* and may be particularly useful for the production of non-pyrogenic about 200 kDa protein, fragments or analogs thereof.

BIOLOGICAL DEPOSITS

Certain plasmids that contain portions and full-length of the gene having the open reading frame of the gene encoding the about 200 kDa outer membrane protein of *M. catarrhalis* strain 4223 that are described and referred to herein have been deposited with the America Type Culture Collection (ATCC) located at 10801 University Blvd., Manassas, VA 20110-2209, U.S.A., pursuant to the Budapest Treaty and pursuant to 37 CFR 1.808 and prior to the filing of this application.

Samples of the deposited plasmids will become available to the public upon grant of a patent based upon this United States patent application or relevant precursor applications. The invention described and claimed herein is not to be limited in scope by plasmids deposited, since the deposited embodiment is intended only as an illustration of the invention. Any equivalent or similar plasmids that encode similar or equivalent antigens as described in this application are within the scope of the invention.

Plasmid	ATCC Designation	Date Deposited
pKS47	97,111	April 7, 1995
pKS5	97,110	April 7, 1995
pKS9	97,114	April 18, 1995
pKS294	203,528	December 17, 1998
pKS348	203,529	December 17, 1998

EXAMPLES

The above disclosure generally describes the present invention. A more complete understanding can be obtained by reference to the following specific Examples. These Examples are described solely for purposes of illustration and are not intended to limit the scope of the invention. Changes in form and substitution of equivalents are contemplated as circumstances may suggest or render expedient. Although specific terms have been employed herein, such terms are

intended in a descriptive sense and not for purposes of limitations.

Methods of molecular genetics, protein biochemistry, and immunology used but not explicitly described in this disclosure and these Examples are amply reported in the scientific literature and are well within the ability of those skilled in the art.

Example 1

This Example describes the cloning of a gene encoding the *M. catarrhalis* 200 kDa outer membrane protein.

A *M. catarrhalis* genomic library in phage lambda EMBL3 was prepared as described in Example 9 of USP 5,808,024 and WO 96/34960 and was screened using guinea pig anti-200 kDa protein antiserum. A lambda phage clone 8II, which expressed an about 200 kDa protein, was confirmed by immunoblotting of the phage lysate using the about 200 kDa outer membrane-specific antiserum.

Plate lysate cultures of this recombinant phage were prepared. The DNA was extracted from the plate lysates using a Wizard Lambda Preps DNA Purification System (Promega Corp, Madison, WI) according to the manufacturer's instructions. This phage clone carried a DNA insert of about 16 kb in size (the restriction map for which is shown in Figure 1). The phage DNA was digested with a mixture of the restriction enzymes *Sal*I and *Xho*I, and separated by agarose gel electrophoresis. Two DNA bands, approximately 5 kb and 11 kb in size, respectively, were cut out from the gel and extracted using a Geneclean kit (BIO 101 Inc., LaJolla, CA) according to the manufacturer's direction.

The smaller 5 kb fragment was ligated into a plasmid vector, pBluescript II SK +/- (Stratagene Cloning Systems, LaJolla, CA), which had been previously digested with *Sal*I and *Xho*I, to produce plasmid pKS5. The larger 11 kb fragment was ligated into a plasmid vector, pSP72 (Promega Corp., Madison, WI), digested

with *Sal*I and *Xho*I,, to produce plasmid pKS9. Both ligated plasmids were used to transform *E. coli*, strain DH5α.

The lambda phage DNA was also digested with a mixture of *Xho*I and *Kpn*I and the approximately 1.1 kb fragment was isolated after agarose gel separation as described above. This 1.1 kb fragment was ligated into a plasmid vector, pGEM-7Zf(+) (Promega Corp., Madison, WI), to produce plasmid pKS47.

Example 2

This Example describes the isolation of chromosomal DNA from *M. catarrhalis* for use in PCR amplification.

M. catarrhalis was cultured in 25 ml of BHI broth overnight and centrifuged at 5,000 rpm for 10 min. The bacteria pellet was suspended in 10 ml of 10 mM Tris/HC1 (pH 8.0) containing 100 mM EDTA and mixed with RNaseA (final concentration: 100 μg/ml) and lysozyme (final concentration: 1 mg/ml). After incubation on ice for 10 min and at room temperature for 50 min, the suspension was gently mixed with 1 ml of 10% SDS and then heated at 65°C for 20 min. The suspension was mixed with proteinase K (final concentration: 200 μg/ml) and incubated at 50°C for 1 h. The suspension was gently mixed with 10 ml chloroform on a nutator for 15 min and centrifuged at 5,000 rpm for 10 min. The upper phase was slowly removed with a wide-bore pipette and mixed with 10 ml of Tris-saturated phenol and 10 ml of chloroform on a nutator. After centrifugation at 5,000 rpm for 10 min, the upper phase was re-extracted with a mixture of Tris-saturated phenol and chloroform, again, and then extracted with chloroform, and then twice dialyzed against 1M NaCl at 4°C and twice against TE buffer (pH 8.0) at 4°C.

Example 3

This Example describes subcloning and sequence analysis of fragments of the 200 kDa protein gene from *M. catarrhalis* strain 4223.

The procedures used to produce a phage λEMBL3 clone 8II, and its subclones, pKS5, pKS9 and pKS47, are described in USP 5,808,024 and WO 96/34960. pKS10 was constructed from the λEMBL3 clone 8II exactly as described for pKS9. pKS59 and pKS63 were constructed by insertion of a 1.4 kb *Xba*I-*Nco*I fragment of pKS9 into pGEM5Z(+) that had been digested with *Nco*I and *Spe*I. pKS71 was made by insertion of the same 1.4 kb *Xba*I-*Nco*I fragment, isolated from the λEMBL3 clone 8II into pGEM5Z(+). Sequence analysis confirmed that all three plasmids, pKS59, pKS63 and pKS71, carried identical DNA fragments. Figure 1 shows partial restriction maps for the plasmids.

The full sequence of the 200 kDa gene locus from the λDNA clone was described in USP 5,808,024 and WO 96/34960 and is shown in Figure 2. There is a tract of 10 consecutive G nucleotides between position 623 and 632 in clones derived from the λ library. The first possible start codon is, therefore, located at nucleotides 706 to 708 and is a GTG encoding a valine, boxed lightly in Figure 2. A series of strains expressing a 200 kDa gene, were identified by immunoblot analysis and the 5' end of their 200 kDa genes was PCR amplified and sequenced. A summary of the findings is shown in Table 5 wherein the expression level of the gene appeared to be related to the number of G nucleotides in the tract and for those strains within higher expression levels, the start codon was an ATG upstream of the GTG codon identified from the 4223 λ clones. Based upon these findings, the sequence of the 5' end of the 200 kDa gene from strain 4223 was re-examined.

Plasmids pKS9 and pKS10 were directly derived from the λ clone. The subclones pKS59 and pKS63 were derived from pKS9 whereas pKS71 contained the same fragment derived directly from the λ clone. All of these plasmids contained 10 G nucleotides in the G tract, as described previously. To determine whether the λ clone contained an extra G nucleotide or the strain itself contained an aberrant gene, PCR amplification of the region was performed from chromosomal DNA preparations and from the λ subclones. The data in Table 3 show that PCR fragments of the λ subclones all contained 10 G nucleotides. The data in Table 4, however, demonstrate that PCR fragments derived directly from chromosomal DNA, contain 9 G nucleotides in the tract. When the single extra G nucleotide is removed from the 200 kDa sequence of strain 4223, the open reading frame is extended in the 5' direction to start from an ATG codon 156 nucleotides earlier, at positions 541 to 543 in Figure 2. This new start codon corresponds to that suggested for the 200 kDa genes sequenced from other strains and summarized in Table 5.

Example 4

This Example describes the construction of the full length 200 kDa protein gene from *M. catarrhalis* strain 4223. The construction scheme is shown in Figure 9.

The full-length 200 kDa protein gene was constructed from the new ATG start codon identified by analysis of the chromosomally derived DNA as described in Example 3 and shown in Figure 3. pKS47 was digested with *Xho*I and *Kpn*I and separated by agarose gel electrophoresis. The 1.1 kb fragment was isolated from the gel and inserted into pKS5, which had previously been digested with the same two enzymes and purified to form pKS80. An about 5.8 kb *Pst*I fragment from pKS80 was inserted into pT7-7 vector (ref. 28) that had been digested with *Pst*I and dephosphorylated. The orientation

of the insert was determined by restriction enzyme analysis and pKS122 was chosen for further construction (see Figure 7).

The 5' region of the 200 kDa protein gene was amplified from strain 4223 chromosomal DNA. PCR reactions were performed using Taq Plus or Tsg Plus enzyme (Sangon Ltd., Scarborough, Ont., Canada) and a Perkin Elmer DNA Thermocycler (Perkin Elmer Cetus, Foster City, CA, USA). The lower PCR reaction mixture (50 μl) contained 5 μl of 10X buffer, 0.4 mM each of four deoxynucleotide triphosphates (Perkin Elmer, Foster City, CA, USA) and 1 to 2 μM each of two primers. The upper PCR reaction mixture (50 μM) contained 5 μl of 10X buffer, 0.5 to 1 μl of Taq Plus or Tsg Plus enzyme, and template DNA. The lower and upper mixtures were separated by a layer of AmpliWax PCR Gem50 (Perkin Elmer, Foster City, CA, USA) before heating cycles started. The thermocycling condition employed for the provision of PCR products in the construction of various plasmids are set forth in Table 11 below. The PCR products were purified using a QIAquick PCR purification kit (Qiagen Inc., Mississauga, Ont., Canada). The purified PCR products were sequenced on both strands directly and/or after cloning in appropriate vectors using an Applied Biosystem sequencer.

The 5' primer (designated 5295.KS) was designed, so that it contained the first possible translation start codon, ATG, and its flanking sequences with a mutation to introduce an *Nde*I site at the ATG. The 3' primer (designated 4260.KS) was based upon the non-coding strand in the region about 1 kb downstream from the ATG start codon. (The nucleic acid sequences and SEQ ID's of the PCR primers utilized herein are identified in Table 10). The PCR-product was digested with *Nde*I and an approximately 650 bp DNA fragment was gel purified and

inserted into pKS122, which had previously been linearized with *Nde*I and dephosphorylated.

The new construct, designated pKS294 (Figure 8), was confirmed by restriction enzyme analyses and by sequencing of the PCR-amplified DNA and its joint regions. The number of G nucleotides in the G tract was nine, and the open reading frame continued from the newly found translation start codon, ATG, to the remaining portion of 200 kDa protein gene in pKS122. pKS294, therefore, carried the correct, full-length 200 kDa protein gene from *Moraxella catarrhalis* strain 4223. During construction of pKS294, *E. coli* strain DH5α was used for transformation and plasmid analyses.

<u>Example 5</u>

This Example describes the cloning and sequence analysis of genes encoding the 200 kDa protein from additional *M. catarrhalis* clinical isolates.

A panel of *M. catarrhalis* clinical isolates was analysed by immunoblot with guinea pig anti-200 kDa antibody, as described in USP 5,808,024 and WO 96/34960. From these analyses, it was evident that there is size heterogencity among the 200 kDa proteins from various strains. In order to assess the possible genetic heterogencity, representative strains were chosen for gene cloning. Strain Q8 is a naturally occurring relatively non-clumping strain that produces a 200 kDa protein of about the same size as the 4223-derived protein. Strain LES-1 produces a larger 200 kDa protein. These strains were also selected based upon bactericidal antibody data as illustrated in Table 1. The 200 kDa genes were cloned from these two strains of M. catarrhalis and sequenced.

The nucleotide and derived amino acid sequences of the 200 kDa genes from strains Q8 and LES-1 are shown in Figures 4 and 5 respectively. An alignment of the amino acid sequences with the 4223-derived sequence is shown in Figure 6. As can be seen, the first 68 residues of

the N-terminus are quite conserved, especially between strains 4223 and Q8. In addition, the final 456 residues of the C-terminus are nearly identical among the three strains. The remainder of the sequence has regions of high homology and significant diversity, including an insert of more than 300 residues for strain LES-1.

The N-terminal sequence of the 200 kDa proteins is homologous to the *H. influenzae* Hia and Hsf proteins, as well as other high molecular weight proteins or adhesins, such as AIDA (ref. 33).

The C-terminal region also has some homology to *H. influenzae* Hia and Hsf proteins as do some stretches of internal sequence. There is also some homology in the C-terminal region to UspA (ref. 23). A further indication of the relatedness of this family of proteins, is the finding that guinea pig anti-200 kDa antibody raised to gel-purified native protein was able to recognize recombinant Hia protein by immunoblot. This data has been described in copending United States Patent Application No. 09/268,347 (Hia) filed March 16, 1999, assigned to the assignee hereof and the disclosure of which is incorporated herein by reference.

<u>Example 6</u>

This Example shows the expression of the full-length about 200 kDa protein from pKS294.

E. coli strain, BL21(DE3)/pLysS was transformed by electroporation with pKS294, prepared as described in Example 4, for the expression study of the full-length 200 kDa protein gene.

The product of the pKS294 construct was found to be toxic to the host *E. coli*. At room temperature, the BL21(DE3)/pLysS transformants grew very slowly on LB-agar plates containing ampicillin (Amp) and chloramphenicol (Cm) and at 37°C, no transformants were detected. When the transformants which grew at room temperature, were cultured overnight at 30°C on BHI agar containing the two antibiotics and glucose, they grew

well, producing colonies with a normal size. However, when these clones were cultured overnight in liquid medium at 30°C, subcultured into broth without glucose, and then induced by addition of IPTG, no recombinant protein was found on Western blot using anti-200 kDa protein serum. When the cells cultured overnight were examined before subculturing, a small quantity of recombinant 200 kDa protein was detected by SDS-PAGE stained with Coomassie Blue and by Western blot, showing that the gene was expressed during the overnight culture.

When *E. coli* strain, DH5α, which cannot express the gene under the control of a T7 promoter, was transformed with pKS294, the transformants grew well at 37°C both on LB-agar and in LB-broth containing the antibiotics. These results suggest that the gene product is very toxic to host *E. coli*, and that even a basal level of expression of the full-length 200 kDa protein gene from the ATG is lethal to *E. coli*.

M. catarrhalis strain LES-1 also produced similar toxicity in *E. coli* when the full length 200 kDa protein was expressed.

<u>Example 7</u>

This Example describes the deletion of a short 5'-sequence from the strain 4223 or strain LES-1 200 kDa protein gene and expression of the truncated genes producing a M56 r200 kDa product.

The deletion of a short 5' region from the 4223 200 kDa protein gene is shown in Figure 10 and was performed using a similar approach as described in Example 4. An about 500 bp 5' region of the 200 kDa gene was PCR amplified from strain 4223 using primers 5471.KS and 4257.KS (Table 8) from chromosomal DNA. The 5' primer (designated 5471.KS) was based upon the region surrounding the previously identified GTG downstream start codon. In primer 5471.KS, the flanking regions around the GTG codon were incorporated and the GTG was

mutated to ATG with further mutations used to introduce an *Nde*I site incorporating the new ATG. Using numbering from the full-length 200 kDa protein, the new start codon would be M56 replacing the previous V56 codon. The 3' primer (designated 4257.KS) was based upon the non-coding strand located about 500 bp downstream from the GTG codon in the 200 kDa protein gene. The PCR-product was digested with *Nde*I, purified using a QIAquick PCR purification kit (Qiagen Inc., Mississauga, Ont.), and inserted into *Nde*I digested and dephosphorylated pKS122 to provide pKS348 (see Figure 7). Plasmid pKS348 was confirmed by restriction enzyme analyses and by sequencing of the PCR-amplified DNA piece and its joint regions. The nucleotide sequence (SEQ ID No: 12) and the deduced amino acid sequence (SEQ ID No: 13) for the 5'-truncation contained in pKS348 are shown in Figure 8. A similar N-terminal truncated 200 kDa gene from strain LES-1 was generated in the same manner and was designated pKS444.

A single colony of *E. coli*, BL21(DE3)/pLysS, (KS358) which carried pKS348, was suspended in 5 ml of BHI broth containing Amp (100 μM), Cm (50 μM) and 0.4% of glucose, and cultured overnight at 37°C. To study the kinetics of expression, 2.5 ml of the overnight culture was added to 250 ml of LB (Luria-Bertani) broth containing Amp (100 μM) and Cm (50 μM), and grown with shaking at 37°C to A_{600} = 0.33 to 0.36. Another 0.3 ml of the overnight culture was added to 30 mL of LB broth containing Amp (100 μM) and Cm (50 μM) and grown with shaking at 37°C to A_{600} = 0.26 to 0.44. Gene expression from the cultures was induced by addition of IPTG (final concentration: 4 mM). The bacteria were grown and harvested at different time points by centrifugation. The expression of the 200 kDa protein gene in the culture was confirmed by SDS-PAGE analysis using Coomassie Blue staining and by Western blot analysis

using guinea pig anti-200 kDa protein serum, as described in USP 5,808,024 and WO 96/34960.

When *E. coli* BL21(DE3)/pLysS was transformed with pKS348, transformants grew well even on LB agar plates and in LB broth containing antibiotics at 37°C. After induction with IPTG, these clones produced a large amount of the N-terminally truncated r200 kDa protein which was clearly seen by SDS-PAGE Coomassie Blue stain, as shown in Figure 12.

The bacterial culture induced at A_{600} = 0.26 produced slightly more truncated r200 kDa protein than the culture induced when the OD reading was 0.44. The largest amount of truncated r200 kDa protein was seen at 3 hr after induction. Similar results were observed for the M56 r200 kDa expression from strain LES-1.

<u>Example 8</u>

This Example describes the purification of the M56 r200 kDa proteins from strain 4223 or LES-1, according to the procedure shown in Figure 11.

E. coli cell pellets were obtained from 500 ml culture prepared as described in Example 7, by centrifugation and were resuspended in 50 ml of 50 mM Tris-HC1, pH 8.0, containing 0.1 M NaCl, and disrupted by sonication. The sonicate was centrifuged at 20,000 xg for 30 min. and the resultant supernatant (sup1) was discarded. The pellet (ppt1) was extracted, in 50 ml of 50 mM Tris-HC1, pH 8.0 containing 0.5% Triton X-100 and 10 mM EDTA, then centrifuged at 20,000 xg for 30 min. and the supernatant (sup2) was discarded. The pellet (ppt2) was further extracted in 50 ml of 50 mM Tris-HC1, pH 8.0, containing 1% octylglucoside, then centrifuged at 20,000 xg for 30 min. and the supernatant (sup3) was discarded.

The resultant pellet (ppt3) contained the inclusion bodies. The pellet was solubilized in 6 ml of 50 mM Tris-HC1, pH 8.0, containing 6 M guanidine and 5 mM DTT. Twelve ml of 50 mM Tris-HCl, pH 8.0 was added, the

mixture centrifuged at 20,000 xg for 30 min, and the pellet (ppt4) discarded. The supernatant (sup4) was precipitated by adding polyethylene glycol (PEG) 4000 at a final concentration of 5% and incubated at 4°C for 30 min. The resultant pellet (ppt5) was removed by centrifugation at 20,000 xg for 30 min. The supernatant was then precipitated by $(NH_4)_2SO_4$ at 50% saturation at 4°C overnight. After the addition of $(NH_4)_2SO_4$, the solution underwent phase separation with protein going to the upper phase (as judged by the cloudiness of the layer). The upper phase was collected, then subjected to centrifugation at 20,000 xg for 30 min. The resultant pellet was collected and dissolved in 2 ml of 50 mM Tris-HCl, pH 8.0, containing 6 M guanidine and 5 mM DTT. The clear solution was purified on a Superdex 200 gel filtration column equilibrated in 50 mM Tris-HC1, pH 8.0, containing 2 M guanidine HC1. The fractions were analysed by SDS-PAGE and those containing the purified r200 kDa were pooled. The pooled fraction was concentrated 5 to 10 fold using a centriprep 30 and then dialysed overnight at 4°C against PBS, and centrifuged at 20,000 xg for 30 min to clarify.

The protein remained soluble under these conditions and glycerol was added to the M56 r200 kDa preparation at a final concentration of 20% for storage at -20°C (Figure 12). The average yield of the purified M56 r200 kDa protein is about 10 mg L^{-1} culture. The purified protein was used for the immunization of animals, as described below.

The procedure of this Example 8 and was repeated for *M. catarrhalis* strain LES-1 and a corresponding r200 kDa protein was produced. The N-terminal truncated M56 r200 kDa protein from strain LES-1 gave approximately the same recovery of purified protein as described above for strain 4223.

Example 9

This Example illustrates the immunogenicity of the M56 r200 kDa protein.

The immunogenicity of M56 r200 kDa, prepared as described in Example 8, was examined using mice and guinea pigs. Groups of five BALB/c mice (Charles River, Quebec) were immunized sub-cutaneously (s.c.) on days 1, 29 and 43 with 0.3, 1.3 and 10 μg of 4223 M56 r200 kDa antigen, prepared as described in Example 8, in the presence $A1PO_4$ (1.5 mg per dose). Blood samples were collected on days 0, 14, 28, 42 and 56.

Groups of five guinea pigs (Charles River, Quebec) were immunized i.m. on days 1, 29 and 43 with 25, 50 and 100 μg of 4223 M56 r200 kDa antigen prepared as described in Example 8, in the presence $A1PO_4$ (1.5 mg per dose). Blood samples were collected on days 0, 14, 28, 42 and 56.

Anti-M56 r200 kDa IgG titers were determined by antigen-specific enzyme-linked immunosorbent assays (EIAs). Microtiter wells (Nunc-MAXISORP, Nunc, Denmark) were coated with 50 μL of protein antigen 0.2 μg mL^{-1}). The reagents used in the assays were as follows: affinity-purified F(ab')2 fragments of goat anti-mouse IgG (Fc-specific) conjugated to horseradish peroxidase (Jackson ImmunoResearch Labs, Mississauga, Ontario); affinity-purified guinea pig anti-IgG antibody (1 μg ml^{-1}) (prepared by the inventors); and affinity-purified F(ab')2 fragment of goat anti-guinea pig IgG (H+L) antibodies conjugated to horseradish peroxidase (HRP) (Jackson ImmunoResearch Laboratories) used as a reporter. The reactions were developed using tetramethylbenzidine (TMB/H_2O_2, ADI, Mississauga, Ontario) and absorbancies were measured at 450 nm (using 540 nm as a reference wavelength) in a Flow Multiskan MCC microplate reader (ICN Biomedicals, Mississauga, Ontario). The reactive titer of an antiserum was defined

as the reciprocal of the dilution consistently showing a two-fold increase in absorbance over that obtained with the pre-bleed serum sample.

The mice generated dose-dependent anti-M56 r200 kDa antibody responses, as shown in Figure 14. These results clearly show that the protein remained immunogenic after inclusion bodies extraction, solubilization and purification. Only a slight difference in the antibody titers were found for the higher dose range tested in guinea pigs (Figure 15), indicating that the amount of antigen used was nearly at saturation.

Example 10

This Example describes the generation of hyper-immune sera against the M56 r200 kDa proteins in rabbits and guinea pigs.

To generate hyper-immune sera against M56 r200 kDa proteins, groups of two rabbits and two guinea pigs (Charles River, Quebec) were immunized intramuscularly (i.m.) on day 1 with a 5 μg dose of purified M56 r200 kDa protein, prepared as described in Example 8, emulsified in complete Freund's adjuvant (CFA). Animals were boosted on days 14 and 29 with the same dose of protein emulsified in incomplete Freund's adjuvant (IFA). Blood samples were taken on day 42 for analyzing the anti-M56 r200 kDa antibody titers and bactericidal activities. Anti-r200 kDa IgG titers were determined by antigen-specific enzyme-linked immunosorbent assays (EIAs), as described in Example 9. The results obtained in the two animals using r200 kDa protein from strains 4223 and LES-1 are illustrated in Table 6.

Example 11

This Example describes a bactericidal antibody assay.

The bactericidal antibody activity of guinea pig anti-M56 r200 kDa sera from 4223 or LES-1 protein prepared as described in Example 10 against various

strains of *M. catarrhalis* was estimated using a viability plating assay. Each test strain of *M. catarrhalis* was cultured overnight in brain heart infusion (BHI) broth (Difco Laboratories, Detroit, MI) at 37°C. The overnight culture was subcultured into 10 ml BHI broth, and grown to an absorbance at 578 nm of 0.5. The number of bacteria at A_{578} = 0.5 changes from strain to strain. Therefore, several ten-fold dilutions of each strain were used in order to achieve 100 to 300 colonies per plate for the preimmune serum group. Bacteria were diluted in Veronal buffered saline (VBS, pH 7.6) containing 140 mM NaCl, 93 mM $NaHCO_3$, 2 mM Na-barbiturate, 4 mM barbituric acid, 0.5 mM $MgCl_2.6H_2O$, 0.4 mM $CaCl_2.2H_2O$, and 0.1% bovine serum albumin. Guinea pig anti-M56 r200 kDa serum and pre-immune control serum were heated at 56°C for 30 min. to inactivate endogenous complement. Serum and antiserum were diluted in VBS, and placed on ice.

Twenty-five µl of diluted pre-immune serum or test antiserum were added to the wells of a 96-well Nunclon microtitre plate (Nunc, Roskilde, Denmark). Twenty-five µl of diluted bacterial cells were added to each of the wells. A guinea pig complement (BioWhittaker, Walkerville, MD) was diluted 1:10 in VBS, and 25 µl portions were added to each well. The plates were incubated for 60 min, gently shaking at 70 rpm on a rotary platform. Fifty µl of each reaction mixture were plated onto Mueller Hinton agar plates (Becton-Dickinson, Cockeysville, MD). The plates were incubated at 37°C for 24 hours, and then left at room temperature for a further 24 hours. The number of colonies per plate was counted, and average values of colonies per plate were estimated from duplicate pairs.

When pre-immune serum plates were compared with PBS control plates (no serum), pre-immune serum had no bactericidal effect on the homologous strain 4223.

Therefore, it was assumed that the number of colonies per plate on pre-immune serum plates represented 100% viability for each strain and percent bactericidal killing was calculated as follows:

$$100\% - \left[\frac{\text{average number of colonies per plate in anti-r200 kDa antiserum group x 100}}{\text{average number of colonies per plate in pre-immune serum group}}\right]\%$$

When the bactericidal antibody activity of the 4223 anti-M56 r200 kDa antiserum was examined against the homologous strain (Table 7), 50% killing was observed at a serum dilution between 1/512 and 1/1024, showing that the antiserum raised against M56 r200 kDa protein possesses bactericidal antibody activity. Next, the bactericidal antibody activity of the antiserum was tested at a dilution of 1/64 against a total of 55 different strains, which were isolated from otitis media patients in various geographical locations (Table 1B). The antiserum raised against the M56 r200 kDa protein from strain 4223 showed more than 30% bactericidal antibody activity against 38 out of 56 (68%) strains examined. When LES-1 anti-M56 r200 kDa antibody was tested in the bactericidal antibody assay, 36/55 (65%) strains were killed, including 11 strains that were not killed by the 4223 anti-M56 r200 kDa antibody. Only six strains out of 55 strains examined were not killed by either one of the two antisera. These results indicate that the 200 kDa protein is a very good candidate for inclusion in an otitis media vaccine.

Example 12

This Example describes the inhibition of binding of *M. catarrhalis* strains to either Chang or Hep-2 epithelial cells by 4223 anti-M56 r200 kDa serum.

The 200 kDa protein had previously been proposed to be an adhesin on the basis of its apparent absence from a spontaneous non-clumping variant of strain 4223. This strain, obtained by serial passaging of culture supernatants, was designated RH408 and is described in WO 96/34960. Electron microcopy also suggested that the

200 kDa protein was an adhesin. The sequence homology demonstrated between the *M. catarrhalis* 200 kDa proteins and other high molecular weight adhesins from different organisms, also suggested that it was an adhesin. Based upon these observations, an assay was developed to try to demonstrate that anti-r200 kDa antibody could block adherence between *M. catarrhalis* and epithelial cells, thus identifying it definitively as an adhesin.

On day 1, 24 well tissue culture plates were seeded with approximately 3 x 10^5 Chang cells per well, to achieve a confluent monolayer following overnight incubation at 37°C in the presence of 5% CO_2. *M. catarrhalis* 4223 or Q8 was cultured in 10 ml of BHI broth at 37°C for 18 hr, shaking at 200 rpm.

On day 2, bacterial cultures were pelleted by centrifugation at 3500 rpm for 10 min, and washed with 10 ml of PBS. After a centrifugation as above, each pellet was resuspended in 2 ml of DMEM supplemented with 10% FBS and 2 mM glutamine. The bacteria cultures were diluted 1/10 in the supplemented DMEM to OD of approximately 1.8 at 578 nm. Confluent monolayers of Chang cells were washed once with 1 ml of PBS per well, and 0.5 ml of 10% BSA in PBS was added to each well as a blocking agent. Plates were incubated at 37°C for 30 min and monolayers were washed twice with PBS as above.

A guinea pig anti-4223 M56 r200 kDa antiserum, prepared as described in Example 10 and pooled pre-immune guinea pig sera were heated at 56°C for 30 min to inactivate endogenous complement. Equal volumes of appropriately diluted antisera and bacteria were mixed, and 200 µl of the mixture were added into each well. Examples of antiserum dilutions tested included 1/4, 1/16 and 1/64. The plate was incubated at 37°C for 1 hr, with gentle shaking. The plate was carefully washed four times with 1 ml of PBS per well to remove the bacteria. To each well, 100 µl of trypsin were added, and the

plate was incubated at 37°C for 5 min. After inactivation of trypsin by addition of 900 μl Dulbecco's Minimal Essential Medium (DMEM) to each well, the cells were resuspended by pipetting up and down several times.

Ten-fold dilutions of resuspended cells were prepared in a new 96-well plate. Fifty μl each of the 1×10^{-2}, 1×10^{-3}, 1×10^{-4} and 1×10^{-5} diluted samples were plated on a Mueller-Hinton agar plate. Plates were incubated at 37°C overnight, and then left at room temperature for a further 24 hours. The number of colonies per plate was counted for the estimation of the total bound bacteria.

Dilution plating was also carried out for each bacterial strain, to estimate bacterial concentrations and to calculate the total amount of bacteria added to each well. It was assumed that the number of bacteria bound to tissue culture cells in the presence of pre-immune sera represented 100% optimal binding for each assay, and 0% inhibition. Therefore, in order to calculate the percent inhibition of the antiserum, we used the following formula:

$$\%\ \text{inhibition} = 100 - \left[\frac{\text{total bacteria bound in 4223 anti-r200 kDa antiserum samples}}{\text{total bacteria bound in pre-immune sera samples}} \times 100\right]$$

When the guinea pig 4223 anti-M56 r200 kDa protein serum was examined for the inhibition of binding of strain 4223 to Chang cells (Table 8), inhibition of 98%, 92% and 83% was observed at antiserum dilutions of 1/4, 1/16 and 1/64, respectively. With the heterologous strain Q8, the inhibition of binding to the tissue culture cells was estimated to be 77%, 82% and 55% at antiserum dilutions of 1/4, 1/16 and 1/64, respectively. The results clearly showed that anti-M56 r200 kDa protein serum inhibited the binding of *M. catarrhalis* to cultured human epithelial cells.

Having demonstrated that 4223 anti-M56 r200 kDa antibody could block adherence of *M. catarrhalis* strains 4223 or Q8 to Chang epithelial cells in a dose-dependent

manner, the studies were extended to other strains. Of particular interest, were those strains that were not killed by anti-M56 r200 kDa antisera in the bactericidal antibody assay. To perform the *in vitro* adherence assay on several strains, a single antibody dilution of 1/16 was used. The data for inhibition of *in vitro* adherence to Hep-2 cells is summarized in Table 9. The procedure for the Hep-2 epithelial cells was identical to the Chang cell procedure described above. The 4223 anti-M56 r200 kDa antibody effectively blocked adherence of the homologous strain by 48%. Strain RH408 does not express the 200 kDa gene and in the assay, antibody inhibited adherence of RH408 to 9%. This would be assumed to be a background level. Of 20 strains tested, 16 were inhibited at rates higher than 9%. Among these strains were 19 strains that had not been killed by the 4223 anti-M56 r200 kDa antibody.

To summarize and as shown in Tables 1, 8 and 9, in our collection of 89 strains of *Moraxella catarrhalis*, 80 express 200 kDa. Of 57 strains tested with 4223 anti-M56 r200 kDa antibody in the bactericidal antibody assay, 39 were killed (58%). An additional 15 strains were inhibited from binding to epithelial cells by the same antibody for a total of 54 strains (95%), against which a single antibody was effective. These data demonstrate the very high potential of r200 kDa proteins as vaccine antigens.

Example 13

This Example describes the sequence analysis of the 200 kDa protein gene from *M. catarrhalis* strain RH408, the non-clumping variant of 4223 described in WO 96/34960.

As described in Example 4 and Table 5, it appeared that the number of G nucleotides in the G tract had a regulatory function on the expression of the 200 kDa gene. *M. catarrhalis* strain 4223 and its non-clumping derivative RH408 appeared to differ only in the

expression of the 200 kDa gene. The 200 kDa gene from strain RH408 was subcloned and sequenced and its sequence compared to the parental gene from strain 4223.

Four partially overlapping fragments of the 200 kDa protein gene were PCR amplified from strain *M. catarrhalis* RH408, using primers illustrated in Figure 16 and Table 10, under the conditions set out in Table 11. The combined sequences of the four PCR products covered approximately 6.5 kb including the entire 200 kDa protein gene and its flanking regions. When the sequence of the 6.5 kb fragment was compared with the sequence of the same region from its parent strain 4223, the only difference was the number of G nucleotides in the G tract. As described in Example 4, the correct number of G nucleotides in the G tract was nine. However, the number G nucleotides in the G tract of RH408 was only eight.

This result, along with the analysis of this region in 24 other strains of *M. catarrhalis* (Table 5) strongly suggests that the number of G nucleotides in the G tract controls the expression of the 200 kDa gene in *M. catarrhalis* strains. Similar mechanisms of transcriptional control are found for other bacterial genes, such as the *N. gonorrhehoae Pilc* gene (ref. 32).

Example 14

This Example describes the generation of additional N-terminal truncated r200 kDa proteins and expression studies.

As described in Example 6, the full-length r200 kDa protein appeared to be toxic to *E. coli* and could not be expressed under normal induction conditions. The M56 r200 kDa proteins were readily expressed, as described in Example 7, and were subsequently shown to be highly promising vaccine candidates in *in vitro* assays (Examples 11 and 12). The expression of r200 kDa proteins of intermediate length and their properties was studied.

Three additional N-terminal truncated 200 kDa genes were constructed from the 4223 200 kDa gene using the procedures described in Example 7. The sites of truncation were chosen based upon and are illustrated in Figure 17. The arrows in Figure 17 indicate the sites of truncation, namely ALA^{12}, VAL^{19} and GLY^{39}, each modified to MET. A 5' fragment up to an internal site was PCR amplified using primers illustrated in Table 8. For the ALA^{12} truncation, the primers were 5′ 6242.ks and 3′ 4257.ks, for the VAL^{19} truncation, the primers were 5′ 6243.ks and 3′ 4257.ks and for the $GL4^{39}$ truncation, the primers were 5′ 6244.ks and 3′ 4257.ks (Table 10). The amplification conditions were the same as those used for pKS348 (Table 11). The PCR products were restricted with *Nde*I and ligated into the *Nde*I sites of pKS348 for expression. While some expression of r200 kDa was obtained with each of the N-terminal truncations, the level did not approach the levels obtained using pKS348.

<u>Example 15</u>

This Example illustrates the construction of plasmids pQWE and pQWF expressing C-terminal fragments of the 200 kDa gene.

As shown in the amino acid comparison of Figure 6, the carboxy half of the 200 kDa protein in quite conserved, the main difference being a large approximately 300 amino acid residue insert in strain LES-1. Since so much cross-reactivity for the anti-M56 r200 kDa antisera had been observed, the conserved carboxy half of the protein was expressed.

Plasmid pKS348 prepared as described in Example 7 was digested with restriction enzymes, *Nde* I and *Nae* I, producing four fragments. The approximately 5.8 kb *Nde* I/*Nae* I fragment containing the T7 promoter, ampicillin antibiotic resistance marker and the 3' end of the 200 kDa gene was agarose gel purified. The approximately 480 bp *Nde* I/*Nde* I fragment containing the 5' end of the 200 kDa gene was also gel purified. This approximately

480 bp fragment was then restriction digested with the enzymes *Nla* IV and *Pst* I and the *Nde* I/*Nla* IV fragment ligated to the previously isolated 5.8 kb *Nde* I/*Nae* I fragment to produce plasmid pQWE, as illustrated in Figure 19. This plasmid construct contained a 200 kDa gene with the *Nla* IV to *Nae* I fragment deleted. This plasmid construct resulted, upon expression as described in Example 7, in a fusion 200 kDa protein containing a very short piece of the 5' end and the 3' half of the 200 kDa protein.

An approximately 500 bp fragment around the *Eco* RI site in the 200 kDa gene from plasmid pKS348 was PCR amplified utilizing a 5' oligonucleotide, 6425.KS and a 3' oligonucleotide 4272.KS (Table 10) using the conditions outlined in Table 11. The 5' oligonucleotide was synthesized with an ATG translational start codon and a *Nde* I restriction site, while the 3' oligonucleotide was synthesized with an *Eco* RI site. The approximately 500 bp PCR fragment was the restriction digested with the enzymes *Nde* I and *Eco* RI. Plasmid pQWE, prepared as described above, was restriction digested with *Nde* I and *Eco* RI as illustrated in Figure 20, and this larger fragment agarose gel purified. The *Nde* I/*Eco* RI PCR fragment was then ligated into the isolated *Nde* I/*Eco* RI fragment from pQWE, to produce plasmid pQWF. This construct expresses a 5' truncated 200 kDa protein, having only the 3' half of this protein from the region about 40 bp upstream of the *Nde* I site to the 3' end.

The constructs pQWE and pQWF, prepared as described above and as illustrated in Figures 19 and 20, were expressed in *E. coli* strain BL21(DE3)/pLysS as described in Example 7. The C-terminal half proteins were obtained at levels of expression approximately twice those achieved using pKS348. Corresponding constructs were prepared from strain LES-1 and produced comparable results.

Antiserum was raised against the C-terminal half of 200 kDa protein produced from construct pQWE following the procedure of Example 10 and was employed in the bactericidal assay described in Example 11. As may be seen in Table 1B the antiserum showed more than 30% of killing against 30 out of 31 strains which were killed by the bactericidal assay using antiserum raised against the product from pKS348.

SUMMARY OF THE DISCLOSURE

In summary of this disclosure, nucleotide sequences encoding an about 200 kDa outer membrane protein from several strains of *Moraxella catarrhalis* are described along with recombinant production of such protein. Modifications are possible within the scope of this invention.

Table 1A

Examination of 200 kDa protein in *M. catarrhalis* strains

STRAIN	ANATOMICAL ORIGIN	SOURCE	EXPRESSION OF 200 kDa PROTEIN
4223	MID. EAR FLUID	T.F. MURPHY	+++
RH408	MUTANT OF 4223		-
3	SPUTUM	"	-
56	SPUTUM	"	-
135	MID. EAR FLUID	'	+++
585	BACTEREMIA	"	-
5191	MID. EAR FLUID	"	+++
8185	NASOPHARYNX	“	+++
M2	SPUTUM	“	+++
M5	SPUTUM	"	-
ATCC25240		ATCC	-
H-04	OTITIS	G.D. CAMPBELL	+++
H-12	“	“	-
PO-34	“	“	+++
PO-51	“	“	+++
E-07	“	“	+++
E-22	“	“	+++
E-23	“	“	+++
E-24	“	“	+++
M-02	“	“	+++
M-20	“	“	+++
M-29	“	“	+++
M-32	“	“	+++
M-35	“	“	+++
Q-2	EXPECTORATION	M.G. BERGERON	+
Q-6	“	“	-
Q-8	“	“	+++
Q-9	“	“	-
Q-10	“	“	+++
Q-11	“	“	+++
Q-12	“	“	-
R-1	BRONCHIAL SECRETIONS	“	+
R-2	“	“	-
R-4	OTITIS	“	+++
R-5	“	“	+++
R-6	“	“	+++
R-7	“	“	+++
N-209	BLOOD	‘	+++
VH-1	OTITIS	V. HOWIE	+++
VH-2	“	“	+++
VH-3	“	“	+++
VH-4	“	“	+++
VH-5	“	“	+++
VH-6	“	“	+++
VH-7	“	“	+++

VH-8	“	“	+++
VH-9	“	“	+++
VH-10	“	“	+++
VH-11	“	“	+++
VH-12	“	“	+++
VH-13	“	“	+++
VH-14	“	“	+++
VH-15	“	“	+++
VH-16	“	“	+++
VH-17	“	“	+++
VH-18	“	“	+++
VH-19	“	“	+++
VH-20	“	“	+++
VH-23	“	“	+++
VH-24	“	“	+++
VH-25	“	“	+++
VH-26	“	“	+++
VH-27	“	“	+++
VH-28	“	“	+++
VH-29	“	“	+++
VH-30	“	“	+++
LES1	OTITIS	L.S. STENFORS	+++
LES2	“	“	+++
LES4	“	“	+++
LES5	“	“	+++
LES6	“	“	+++
LES7	“	“	+++
LES9	“	“	+++
LES9	“	“	+++
LES10	“	“	+++
LES11	“	“	+++
LES12	“	“	+++
LES13	“	“	+++
LES16	“	“	+++
LES17	“	“	+++
LES21	“	“	+++
30607	OTITIS	C.W. FORD	+++
CJ1	“	C. JOHNSON	+++
CJ3	“	“	+++
CJ4	“	“	+++
CJ7	“	“	+++
CJ8	“	“	+++
CJ9	“	“	+++
CJ11	“	“	+++

Bacteria were lysed and proteins were separated on SDS-PAGE gels. The expression of 200 kDa protein was examined by Coomassie Blue staining and by Western blot using anti-200 kDa protein guinea pig serum.

TABLE 1B

Bactericidal assay results against *Moraxella catarrhalis* using antisera raised against recombinant M56 200 kDa protein from strains 4223 and LES1, and recombinant C-terminal half of 200 kDa protein from strain 4223.

STRAIN	Killed by anti-M56 200 kDa from 4223	Killed by anti-C-terminal half of 200 kDa from 4223	Killed by anti-M56 200 kDa from LES1
4223	++	++	-
135	++	++	++
H-04	++	++	?
H-12*	-	NT	-
PO-34	-	NT	++
PO-51	-	NT	-
E-07	-	NT	++
E-22	++	++	-
E-24	-	NT	-
M-02	++	++	++
M-20	++	+	-
M-29	++	++	++
M-32	++	++	++
M-35	++	++	++
R4	-	NT	++
R5	++	++	++
R6	++	+	+
R7	++	NT	?
Q8**	++	+	NT
VH-1	++	NT	++
VH-2	++	NT	++
VH-4	-	NT	++
VH-5	++	++	-
VH-7	++	+	?
VH-8	++	++	++
VH-9	-	NT	++
VH-10	++	++	++
VH-13	-	NT	-
VH-15	++	++	++
VH-17	-	NT	-
VH-19	++	++	++
VH-20	+	+	++
VH-23	+	NT	++
VH-24	++	++	-
VH-25	-	NT	++
VH-26	-	NT	++
VH-27	-	NT	-
VH-28	+	NT	-
VH-29	++	++	++
VH-30	-	NT	++
LES1	-	NT	++
LES2	++	++	+
LES4	+	NT	++
LES5	-	NT	++
LES9	++	++	++
LES11	+	+	+
LES12	-	NT	?
LES13	-	NT	++
LES16	+	++	++
LES17	++	++	-
LES21	++	++	-
30607	+	NT	++

CJ1	++	++	++
CJ3	++	-	++
CJ4	++	++	++
CJ7	++	++	++
CJ8	++	++	?

* This strain does not produce 200 kDa protein.

** This is the only non-otitis media strain (isolated from expectorate) in this Table.

++: Killed more than 60% (>60%), +: killed between 30% and 60%, -: killed 30% or less, NT: not tested, ?: the results not tested.

TABLE 2

The number of G nucleotides in the G tract of the 200 kDa protein gene determined by sequencing of subcloned genes from a λEMBL3 clone.

Plasmid*	Number of G's
pKS10	10
pKS59	10
PKS63	10
PKS71	10

* pKS10 and pKS71 carried a DNA insert directly subcloned from a λEMBL3 clone. pKS59 and pKS63 carried a subcloned DNA fragment, pKS9, which was a subclone from an λEMBL3 clone. pKS59, pKS63 and pKS71 carried identical DNA inserts.

TABLE 3

The number of G nucleotides in the G tract of the 200 kDa protein gene amplified by PCR from subcloned genes

Primers	Template DNA	Number of G's
4211 and 4213	pKS9	10
4211 and 4213	pKS10	10
4211 and 4213	pKS71	10

* pKS9, pKS10 and pKS71, which contain a 5' fragment of the 200 kDa protein gene, were independently subcloned from the λEMBL3 clone.

TABLE 4

The number of G nucleotides in the G tract of the 200 kDa protein gene amplified by PCR from chromosomal DNA of strain 4223

Primers	Template	Number of G
4211 and 4166	4223B	9
4211 and 4213	4223B	9
4211 and 4213	4223R	9

* The template chromosomal DNAs, 4223B and 4223R, were independently prepared from *M. catarrhalis* strain 4223.

TABLE 5

The number of G nucleotides in the G tract in different strains of *M. catarrhalis*

Expression	Number of G	Number of strains examined	Possible start codon
+++	3	1	ATG
+++	6	7	ATG
+++	9	7	ATG
+	10	3	GTG
-	7	3	GTG
-	8	2	GTG
-	9	1*	ATG
Total		24	

* The 200 kDa protein gene of this strain was prematurely terminated by a stop codon.

TABLE 6

Anti-M56 r200 kDa antibody titers in guinea pig and rabbit sera

ANTISERA	ANTIBODY TITERS	
	Against M56 r200 kDa (4223)	Against M56 r200 kDa (LES-1)
Gp anti-r200 kDa (4223)	204,800 409,600	102,400 409,600
Gp anti-r200 kDa (LES1)	204,800 102,400	1,638,400 1,638,400
Rb anti-r200 kDa (4223)	102,400 102,400	102,400 102,400
Rb anti-r200 kDa (LES1)	25,600 102,400	204,800 409,600

TABLE 7

Killing of *M. catarrhalis* strain 4223 by the bactericidal antibody activity of guinea pig anti-M56 r200 kDa protein serum

Serum dilution	1/64	1/128	1/256	1/512	1/1024
Killing %	97%	95%	95%	80%	38%

* The guinea pig antiserum was raised against M56 r200 kDa protein from strain 4223, and the bactericidal antibody activity of the serum at various dilutions were examined against the strain 4223.

TABLE 8

Inhibition of the binding of *M. catarrhalis* strains to Chang cells by guinea pig anti-M56 r200 kDa protein serum

Strain	1/4	1/16	1/64
4223	98%	92%	83%
Q8	77%	82%	55%

* The guinea pig antiserum was raised against M56 r200 kDa protein from strain 4223.

TABLE 9

Inhibition of *in vitro* adherence of *Moraxella catarrhalis* to Hep-2 cells by antiserum raised against recombinant 200 kDa protein from strain 4223

STRAIN	Inhibition
4223*	+++
PO-34	+++
PO-51	++
E-07	++
R4	++
VH-4	++
VH-9	-
VH-13	+
VH-17	++
VH-23	++
VH-25	++
VH-26	+++
VH-27	+
VH-28	+++
LES1	++
LES4	-
LES12	-
LES13	-
30607	+

+++: Inhibition was 30% or higher, ++: Inhibition was 20% to 30%, +: Inhibition was 15% to 20%, -: Inhibition was lower than 15%.

*: This strain is the positive control, and the only strain in this Table, which was killed by the bactericidal activity of anti-recombinant 200 kDa protein serum.

TABLE 10

Nucleotide sequences of primers used for PCR amplifications

PRIMER	NUCLEOTIDE SEQUENCE	SEQ ID No:
4211.KS	GATGCCTACGAGTTGATTTGGGT	14
4213.KS	GAGCGTTGCACCGATCACGAGGA	15
4166.KS	CACTAGCCTTTACATCACCACCGATG	16
5295.KS	AAGGTAAACCCATATGAATCACATCTATAAAGTCA	17
4260.KS	GCTTCTAGCTGTGCCACATTGA	18
5471.KS	CGCTCGCTGTCCATATGATCGGTGCAACGCTCA	19
4257.KS	GACCCTGTGCATATGACATGGCT	20
4254.KS	CCTTGGCATCAATCGTGGCACA	21
4278.KS	TTACCTGCATCAATGCCATTGTCT	22
4329.KS	CTGAGGTGAATACAACTACA	23
4272.KS	CATCAGAGGTCTTTGAGGTGTCAT	24
4118.KS	CATCACCGTGGGTCAAAAGAACGCA	25
4267.KS	GATGTCGGCAATGTTTACCTGA	26
4269.KS	CCACATTGACCAGTACTGGCACAGGTGCTA	27
4981.KS	ACCTATGATCAATGGCGATTTGGT	28
6425.KS	AAAGATCATATGGTTACCTTTGGCATTAAC	29
6242	GTCATCTTTCATATGGCCACAGGCACA	30
6243	ACATTTATGCATATGGCAGAGTACGCCA	31
6244	GCTACAGGGCATATGGGCAGTGTATGCACT	32

TABLE 11

PCR Cycle Conditions

1. For the construction of pKS294, oligonucleotides 5295 and 4260 and of pKS348, oligonucleotides 5471 and 4257:
   95°C for 2 min → 95°C for 1 min, 60°C for 30 sec, 72°C for 1 min (10 cycles) → 95°C for 1 min, 62°C for 30 sec, 72°C for 1 min (20 cycles with extension of 1 sec/cycle) → 72°C for 10 min → 4°C.
2. For the construction of pQWF, oligonucleotides 6425 and 4272:
   95°C for 2 min → 95°C for 1 min, 60°C for 30 sec, 72°C for 1 min (10 cycles) → 95°C for 1 min, 60°C for 30 sec, 72°C for 1 min (20 cycles with extension of 1 sec/cycle) → 72°C for 10 min → 4°C.
3. For the amplification of 700 bp fragment for sequencing the G-nucleotide tract from different strains, oligonucleotides 4211 and 4166.
   95°C for 2 min → 95°C for 1 min, 60°C for 1 min, 72°C for 2 min (10 cycles) → 95°C for 1 min, 60°C for 1 min, 72°C for 2 min (20 cycles with extension of 5 sec/cycle) → 72°C for 10 min → 4°C.
4. For sequencing 200 kDa protein from *M. catarrhalis* strain RH408,
   (a) oligonucleotides 4254 and 4278; 4118 and 4267; and 4269 and 4981:
   95°C for 2 min → 95°C for 1 min, 62°C for 30 sec, 72°C for 1 min (10 cycles) → 95°C for 1 min, 62°C for 30 sec, 72°C for 1 min (20 cycles with extension of 2 sec/cycle) → 72°C for 10 min → 4°C.
   (b) oligonucleotides 4329 and 4272
   95°C for 2 min → 95°C for 1 min, 58°C for 30 sec, 72°C for 1 min 30 sec (10 cycles) → 95°C for 1 min, 58°C for 30 sec, 72°C for 1 min 30 sec (20 cycles with extension of 1 sec/cycle) → 72°C for 10 min → 4°C.

REFERENCES

1. Van Hare, G.F., P.A. Shurin, C.D. Marchant, N.A. Cartelli, C.E.Johnson, D. Fulton, S. Carlin, and C.H. Kim. Acute otitis media caused by *Branhamella catarrhalis*: biology and therapy.(1987) Rev. Infect. Dis. 9:16-27.

2. Chapman, A.J., D.M. Musher, S. Jonsson, J.E. Clarridge, and R.J. Wallace. 1985. Development of bactericidal antibody during *Branhamella catarrhalis* infection. J. Infect. Dis. 151:878-882.

3. Hager, H., A. Verghese, S. Alvarez, and S.L. Berk. 1987. *Branhamella catarrhalis* respiratory infections. Rev. Infect. Dis. 9:1140-1149.

4. McLeod, D.T., F. Ahmad, M.J. Croughan, and M.A. Calder. 1986. Bronchopulmonary infection due to *M. catarrhalis*. Clinical features and therapeutic response. Drugs 31(Suppl.3):109-112.

5. Nicotra, B., M. Rivera, J.I. Luman, and R.J. Wallace. 1986. *Branhamella catarrhalis* as a lower respiratory tract pathogen in patients with chronic lung disease. Arch.Intern.Med. 146:890-893.

6. Ninane, G., J. Joly, and M. Kraytman. 1978. Bronchopulmonary infection due to *Branhamella catarrhalis* : 11 cases assessed by transtracheal puncture. Br.Med.Jr. 1:276-278.

7. Srinivasan, G., M.J. Raff, W.C. Templeton, S.J. Givens, R.C. Graves, and J.C. Mel. 1981. *Branhamella catarrhalis* pneumonia. Report of two cases and review of the literature. Am.Rev. Respir. Dis. 123:553-555.

8. West, M., S.L. Berk, and J.K. Smith. 1982. *Branhamella catarrhalis* pneumonia. South.Med. J. 75:1021-1023.

9. Brorson, J-E., A. Axelsson, and S.E. Holm. 1976. Studies on *Branhamella catarrhalis* (*Neisseria catarrhalis*) with special reference to maxillary sinusitis. Scan. J. Infect. Dis. 8:151-155.

10. Evans, F.O., Jr., J.B. Sydnor, W.E.C. Moore, G.R. Moore, J.L. Manwaring, A.H. Brill, R.T. Jackson, S. Hanna, J.S. Skaar, L.V. Holdeman, G.S. Fitz-Hugh, M.A. Sande, and J.M. Gwaltney, Jr. 1975. Sinusitis of the maxillary antrum. N.Engl.J.Med. 293:735-739.

11. Tinkelman, D.G., and H.J. Silk. 1989. Clinical and bacteriologic features of chronic sinusitis in children. Am.J.Dis.Child. 143:938-942.

12. Wald, E.R., C. Byers, N.Guerra, M.Casselbrant, and D. Beste. 1989. Subacute sinusitis in children. J.Pediatr. 115:28-32.

13. Wald, E.R., G.J. Milmoe, A. Bowen, J.Ledesma-Medina, N. Salamon, and C.D.Bluestone. 1981. Acute maxillary sinusitis in children. N.Engl.J.Med. 304:749-754.

14. Christensen, J.J., and B. Bruun. 1985. Bacteremia caused by a beta-lactamase producing strain of *Branhamella catarrhalis*. Acta.Pathol. Microbiol. Immunol. Scand. Sect.B 93:273-275.

15. Craig, D.B., and P.A. Wehrle. 1983. *Branhamella catarrhalis* septic arthritis. J. Rheumatol. 10:985-986.

16. Gray, L.D., R.E. Van Scoy, J.P. Anhalt, and P.K.W. Yu. 1989. Wound infection caused by *Branhamella catarrhalis*. J.Clin.Microbiol. 27:818-820.

17. Guthrie, R., K. Bakenhaster, R.Nelson, and R. Woskobnick. 1988. *Branhamella catarrhalis* sepsis: a case report and review of the literature. J.Infect.Dis. 158:907-908.

18. Hiroshi, S., E.J. Anaissie, N.Khardori, and G.P. Bodey. 1988. *Branhamella catarrhalis* septicemia in patients with leukemia. Cancer 61:2315-2317.

19. O'Neill, J.H., and P.W. Mathieson. 1987. Meningitis due to *Branhamella catarrhalis*. Aust. N.Z. J. Med. 17:241-242.

20. Murphy, T.F. 1989. The surface of *Branhamella catarrhalis*: a systematic approach to the surface antigens of an emerging pathogen. Pediatr. Infect. Dis. J. 8:S75-S77.

21. Klingman, K.L., and T.F. Murphy. 1994. Purification and characterization of a high-molecular-weight outer membrane protein of *Moraxella (Branhamella) catarrhalis*. Infect. Immun. 62:1150-1155.

22. Helminen, M.E., I. Maciver, J.L. Latimer, J. Klesney-Tait, L.D. Cope, M. Paris, G.H. McCracken, Jr., and E.J. Hansen. 1994. A large, antigenically conserved protein on the surface of *Moraxella catarrhalis* is a target for protective antibodies. J. Infect. Dis. 170:867-872.

23. Aebi, C., I. Maciver, J.L. Latimer, L.D. Cope, M.K. Stevens, S.E. Thomas, G.H. McCracken, Jr., and E.J. Hansen. 1997. A protective epitope of Moraxella catarrhalis is encoded by two different genes. Infect. Immun. 65:4367-4377.

24. Nixon-George et al. The adjuvant effect of Stearyl Tyrosine on a recombinant subunit hepatitis B surface antigen. (1990), J. Immunology 144:4798-4802.

25. Wiesmuller (1989), Novel low-molecular weight synthetis vaccine against foot-and-mouth disease containing a potent B-cell and macrophage activator, Vaccine 8:29-33.

26. Deres et al. (1989), Nature 342:561.

27. Lockhoff, O. Glycolipids as Immmunomodulators: Synthesis and Properties. 1991. Chem. Int. Ed. Engl. 30:1611-1620.

28. Taber and Richardson, 1985, PNAS 82(4):1074-8.

29. Fitzgerald et al, FEMS Immunol. & Med. Microbiol. 18:209-216, 1997.

30. Fitzgerald et al, FEMS Immunol. & Med Microbiol. 23:57-66, 1999.

31. Kyd et al, J. Med. Microbiology, 47:159-168, 1998.

32. Johson et al, 1997, EMBO J. 10(2):477-488.

33. Benz, I. and Schmidt, M.A., 1992. Mol Microbiol 6:1539-1546.

CLAIMS

What we claim is:

1. An isolated and purified nucleic acid molecule having a nucleotide sequence selected from the group consisting of:

    (a) a nucleotide sequence set forth in Figure 3, 4 or 5 (SEQ ID Nos: 5, 6, 8, 9, 11, 12) for *Moraxella catarrhalis* strains 4223, Q8 and LES-1 respectively or the complementary sequence thereto,

    (b) a nucleotide sequence encoding an about 200 kDa outer membrane protein of a strain of *Moraxella catarrhalis* and having the derived amino acid sequence shown in Figures 3, 4 or 5 (SEQ ID Nos: 7, 10, 13) for *Moraxella catarrhalis* strains 4223, Q8 and LES-1 respectively, and

    (c) a nucleotide sequence encoding an about 200 kDa outer membrane protein of another strain of *Moraxella catarrhalis* which is characterized by a tract of consecutive G nucleotides which is 3 or a multiple thereof in length, an ATG start codon about 80 to 90 bp upstream of said tract and said tract being located between about amino acids 25 and 35 encoded by the nucleotide sequence.

2. The nucleic acid molecule of claim 1 wherein said another strain of *Moraxella catarrhalis* in (c) is a strain as identified in Table 1A other than strains 4223, Q8 and LES-1 and expressing an about 200 kDa protein.

3. An isolated and purified nucleic acid molecule having a nucleotide sequence selected from the group consisting of:

    (a) a nucleotide sequence set forth in Figure 8 (SEQ ID No: 12) for a 5'-truncation of the gene encoding an about 200 kDa outer membrane protein of *Moraxella catarrhalis* strain 4223,

(b) a nucleotide sequence encoding the derived amino acid sequence set forth in Figure 9 (SEQ ID No: 13) for a N-terminal truncation of an about 200 kDa outer membrane protein of *Moraxella catarrhalis* strain 4223, and

(c) a nucleotide sequence encoding a 5'-truncation of a gene encoding an about 200 kDa outer membrane protein of another strain of *Moraxella catarrhalis* and being capable of expressing the corresponding N-terminally truncated about 200 kDa outer membrane protein from *E. coli*.

4. An isolated and purified nucleic acid molecule which is a contiguous *Nde* I - *Pst* I fragment of SEQ ID No: 5.

5. A vector for transforming a host comprising a nucleic acid molecule as claimed in any one of claims 1 to 4.

6. The vector of claim 5 which is a plasmid vector.

7. The vector of claim 5 which has the identifying characteristics of pKS348 (ATCC 203529) shown in Figure 10 or pKS294 (ATCC 203528) shown in Figure 9.

8. The vector of claim 5 which has the identifying characteristic of pQWE shown in Figure 19 or pQWF shown in Figure 20.

9. A host cell transformed by a vector as claimed in claim 5 and expressing an about 200 kDa protein of a strain of *Moraxella catarrhalis* or an approximately C-terminal half thereof.

10. The host cell of claim 9 which is *E. coli*.

11. A recombinant about 200 kDa outer membrane protein of a strain of *Moraxella catarrhalis* or an approximately C-terminal half thereof producible by the transformed host of claim 9.

12. The recombinant protein of claim 11 producible in inclusion bodies.

13. An immunogenic composition comprising the recombinant about 200 kDa outer membrane protein or an approximately C-terminal half thereof of claim 11.

14. The immunogenic composition of claim 13 formulated as a vaccine for *in vivo* administration to protect against disease caused by *Moraxella catarrhalis*.

15. The immunogenic composition of claim 13 in combination with a targeting molecule for delivery to specific cells of the immune system or to mucosal surfaces.

16. The immunogenic composition of claim 13 formulated as a microparticle, capsule or liposome preparation.

17. The immunogenic composition of claim 13 further comprising an adjuvant.

18. A method of inducing protection against disease caused by *Moraxella catarrhalis*, comprising administering to a susceptible host an effective amount of the immunogenic composition of claim 13.

19. The method of claim 18 wherein said susceptible host is a human.

20. A method for the production of an about 200 kDa outer membrane protein of a strain of *Moraxella catarrhalis* or an approximately C-terminal half thereof, which comprises:

transforming a host with a vector as claimed in claim 5,

growing the host cell to express the encoded about 200 kDa protein or an approximately C-terminal half thereof, and

isolating and purifying the expressed about 200 kDa protein or an approximately C-terminal half thereof.

21. The method of claim 20 wherein the host cell is *E. coli*.

22. The method of claim 20 wherein said encoded about 200 kDa protein is expressed in inclusion bodies.

23. The method of claim 22 wherein said isolation and purification of the expressed about 200 kDa protein is effected by:

disrupting the grown transformed cells to produce a supernatant and the inclusion bodies,

solubilizing the inclusion bodies to produce a solution of the recombinant about 200 kDa protein,

chromatographically purifying the solution of recombinant about 200 kDa protein free from contaminating proteins, and

isolating the purified recombinant about 200 kDa protein.

ABSTRACT OF THE DISCLOSURE

An isolated and purified outer membrane protein of a *Moraxella* strain, particularly *M. catarrhalis*, having a molecular mass of about 200 kDa, is provided by recombinant means. The about 200 kDa outer membrane protein as well as nucleic acid molecules encoding the same are useful in diagnostic applications and immunogenic compositions, particularly for *in vivo* administration to a host to confer protection against disease caused by a bacterial pathogen that produces the about 200 kDa outer membrane protein or produces a protein capable of inducing antibodies in a host specifically reactive with the about 200 kDa outer membrane protein.

FIGURE 1

Subclones of portions of the 200 kDa protein gene from λEMBL3 clone 8II and PCR amplification of 5' region

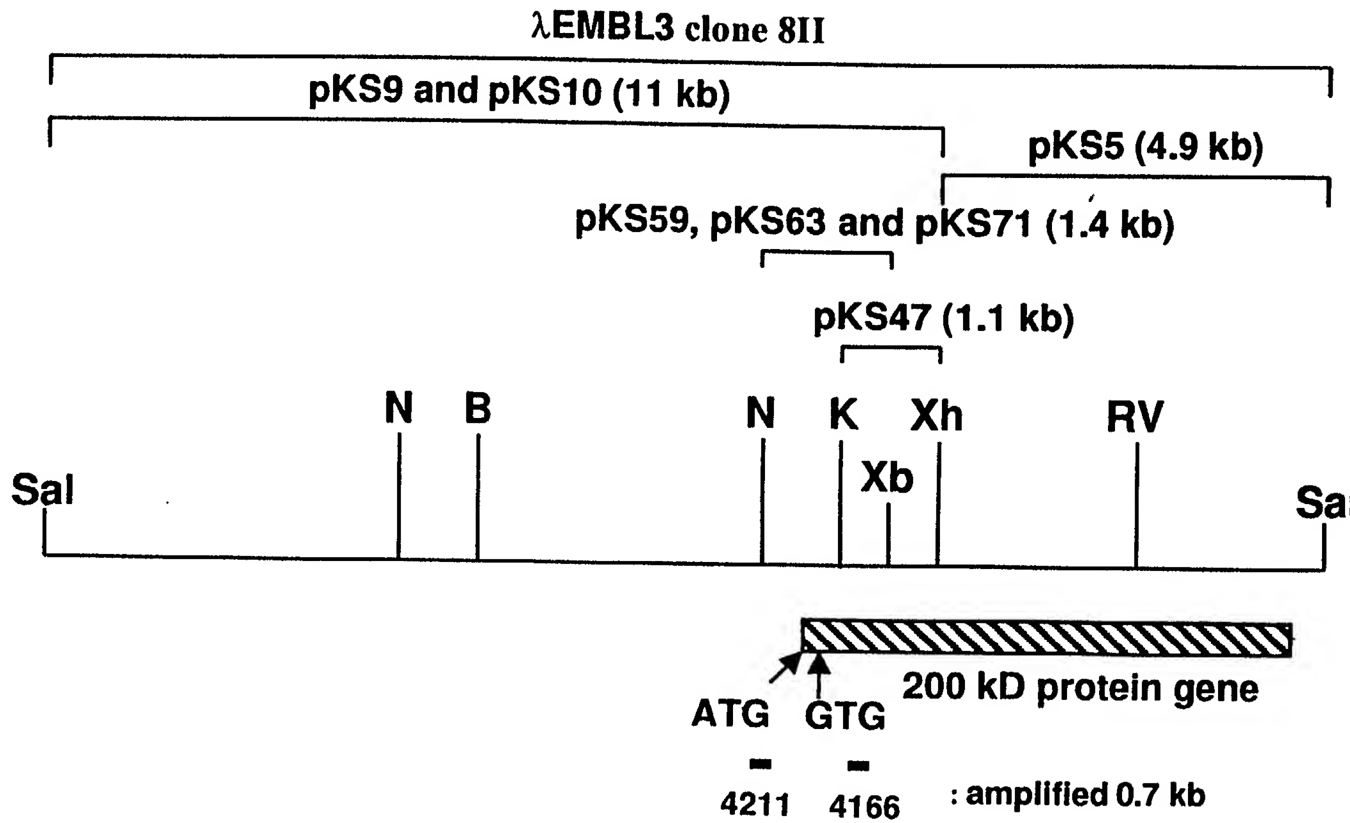


Figure 2. *M. catarrhalis* strain 4223 λEMBL3 clone 200kDa gene

```
ccatggatat gggcaggtgt gctcgcctgc cgtatgatgg cgatgacacc ccatttgccc 60

catatctgta cgatttgaca tgtgatatga tttaacatgt gacatgattt aacattgttt 120

aatactgttg ccatcattac cataatttag taacgcattt agtaacgcat ttgtaaaaat 180

cattgcgccc ctttatgtgt atcatatgaa tagaatatta tgattgtatc tgattattgt 240

atcagaatgg tgatgctata tgatgatgcc tacgagttga tttgggttaa tcactctatg 300

atttgatata ttttgaaact aatctattga cttaaatcac catatggtta taatttagca 360

taatggtagg ctttttgtaa aaatcacatc gcaatattgt tctactgtta ctaccatgct 420

tgaatgacga tcccaatcac cagattcatt caagtgatgt gtttgtatac gcaccattta 480

ccctaattat ttcaatcaaa tgcctatgtc agcatgtatc atttttttaa ggtaaaccac 540

catgaatcac atctataaag tcatctttaa caaagccaca ggcacattta tggcagtggc 600

agagtacgcc aaatcccaca gcacgggggg ggggtagctg tgctacaggg caagttggca 660

gtgtatgcac tctgagcttt gcccgtattg ccgcgctcgc tgtcctc gtg atc ggt    716
                                                    Val Ile Gly
                                                     1

gca acg ctc agt ggc agt gct tat gct caa aaa aaa gat acc aaa cat    764
Ala Thr Leu Ser Gly Ser Ala Tyr Ala Gln Lys Lys Asp Thr Lys His
     5                  10                  15

atc gca att ggt gaa caa aac cag cca aga cgc tca ggc act gcc aag    812
Ile Ala Ile Gly Glu Gln Asn Gln Pro Arg Arg Ser Gly Thr Ala Lys
 20                 25                  30                  35

gcg gac ggt gat cga gcc att gct att ggt gaa aat gct aac gca cag    860
Ala Asp Gly Asp Arg Ala Ile Ala Ile Gly Glu Asn Ala Asn Ala Gln
                40                  45                  50

ggc ggt caa gcc atc gcc atc ggt agt agt aat aaa act gtc aat gga    908
Gly Gly Gln Ala Ile Ala Ile Gly Ser Ser Asn Lys Thr Val Asn Gly
            55                  60                  65

agc agt ttg gat aag ata ggt acc gat gct acg ggt caa gag tcc atc    956
Ser Ser Leu Asp Lys Ile Gly Thr Asp Ala Thr Gly Gln Glu Ser Ile
        70                  75                  80

gcc atc ggt ggt gat gta aag gct agt ggt gat gcc tcg att gcc atc    1004
Ala Ile Gly Gly Asp Val Lys Ala Ser Gly Asp Ala Ser Ile Ala Ile
    85                  90                  95

ggt agt gat gac tta cat ttg ctt gat cag cat ggt aat cct aaa cat    1052
Gly Ser Asp Asp Leu His Leu Leu Asp Gln His Gly Asn Pro Lys His
100                 105                 110                 115

ccg aaa ggt act ctg att aac gat ctt att aac ggc cat gca gta tta    1100
Pro Lys Gly Thr Leu Ile Asn Asp Leu Ile Asn Gly His Ala Val Leu
                120                 125                 130
```

aaa gaa ata cga agc tca aag gat aat gat gta aaa tat aga cgc aca 1148
Lys Glu Ile Arg Ser Ser Lys Asp Asn Asp Val Lys Tyr Arg Arg Thr
135 140 145

acc gca agc gga cac gcc agt act gca gtg gga gcc atg tca tat gca 1196
Thr Ala Ser Gly His Ala Ser Thr Ala Val Gly Ala Met Ser Tyr Ala
150 155 160

cag ggt cat ttt tcc aac gcc ttt ggt aca cgg gca aca gct aaa agt 1244
Gln Gly His Phe Ser Asn Ala Phe Gly Thr Arg Ala Thr Ala Lys Ser
165 170 175

gcc tat tcc ttg gca gtg ggt ctt gcc gcc aca gcc gag ggc caa tct 1292
Ala Tyr Ser Leu Ala Val Gly Leu Ala Ala Thr Ala Glu Gly Gln Ser
180 185 190 195

aca atc gct att ggt tct gat gca aca tct agc tcg ttg gga gcg ata 1340
Thr Ile Ala Ile Gly Ser Asp Ala Thr Ser Ser Ser Leu Gly Ala Ile
200 205 210

gcc ctt ggt gca ggt act cgt gct cag cta cag ggc agt att gcc cta 1388
Ala Leu Gly Ala Gly Thr Arg Ala Gln Leu Gln Gly Ser Ile Ala Leu
215 220 225

ggt caa ggt tct gtt gtc act cag agt gat aat aat tct aga ccg gcc 1436
Gly Gln Gly Ser Val Val Thr Gln Ser Asp Asn Asn Ser Arg Pro Ala
230 235 240

tat aca cca aat acc cag gca cta gac ccc aag ttt caa gcc acc aat 1484
Tyr Thr Pro Asn Thr Gln Ala Leu Asp Pro Lys Phe Gln Ala Thr Asn
245 250 255

aat acg aag gcg ggt cca ctt tcc att ggt agt aac tct atc aaa cgt 1532
Asn Thr Lys Ala Gly Pro Leu Ser Ile Gly Ser Asn Ser Ile Lys Arg
260 265 270 275

aaa atc atc aat gtc ggt gca ggt gtt aat aaa acc gat gcg gtc aat 1580
Lys Ile Ile Asn Val Gly Ala Gly Val Asn Lys Thr Asp Ala Val Asn
280 285 290

gtg gca cag cta gaa gcg gtg gtg aag tgg gct aag gag cgt aga att 1628
Val Ala Gln Leu Glu Ala Val Val Lys Trp Ala Lys Glu Arg Arg Ile
295 300 305

act ttt cag ggt gat gat aac agt act gac gta aaa ata ggt ttg gat 1676
Thr Phe Gln Gly Asp Asp Asn Ser Thr Asp Val Lys Ile Gly Leu Asp
310 315 320

aat act tta act att aaa ggt ggt gca gag acc aac gca tta acc gat 1724
Asn Thr Leu Thr Ile Lys Gly Gly Ala Glu Thr Asn Ala Leu Thr Asp
325 330 335

aat aat atc ggt gtg gta aaa gag gct gat aat agt ggt ctg aaa gtt 1772
Asn Asn Ile Gly Val Val Lys Glu Ala Asp Asn Ser Gly Leu Lys Val
340 345 350 355

aaa ctt gct aaa act tta aac aat ctt act gag gtg aat aca act aca 1820
Lys Leu Ala Lys Thr Leu Asn Asn Leu Thr Glu Val Asn Thr Thr Thr

360 365 370

tta aat gcc aca acc aca gtt aag gta ggt agt agt agt agt act aca 1868
Leu Asn Ala Thr Thr Thr Val Lys Val Gly Ser Ser Ser Ser Thr Thr
375 380 385

gct gaa tta ttg agt gat agt tta acc ttt acc cag ccc aat aca ggc 1916
Ala Glu Leu Leu Ser Asp Ser Leu Thr Phe Thr Gln Pro Asn Thr Gly
390 395 400

agt caa agc aca agc aaa acc gtc tat ggc gtt aat ggg gtg aag ttt 1964
Ser Gln Ser Thr Ser Lys Thr Val Tyr Gly Val Asn Gly Val Lys Phe
405 410 415

act aat aat gca gaa aca aca gca gca atc ggc act act cgt att acc 2012
Thr Asn Asn Ala Glu Thr Thr Ala Ala Ile Gly Thr Thr Arg Ile Thr
420 425 430 435

aga gat aaa att ggc ttt gct cga gat ggt gat gtt gat gaa aaa caa 2060
Arg Asp Lys Ile Gly Phe Ala Arg Asp Gly Asp Val Asp Glu Lys Gln
440 445 450

gca cca tat ttg gat aaa aaa caa ctt aaa gtg ggt agt gtt gca att 2108
Ala Pro Tyr Leu Asp Lys Lys Gln Leu Lys Val Gly Ser Val Ala Ile
455 460 465

acc ata gac aat ggc att gat gca ggt aat aaa aag atc agt aat ctt 2156
Thr Ile Asp Asn Gly Ile Asp Ala Gly Asn Lys Lys Ile Ser Asn Leu
470 475 480

gcc aaa ggt agc agt gct aac gat gcg gtt acc atc gaa cag ctc aaa 2204
Ala Lys Gly Ser Ser Ala Asn Asp Ala Val Thr Ile Glu Gln Leu Lys
485 490 495

gcc gcc aag cct act tta aac gca ggc gct ggc atc agt gtc aca cct 2252
Ala Ala Lys Pro Thr Leu Asn Ala Gly Ala Gly Ile Ser Val Thr Pro
500 505 510 515

act gaa ata tca gtt gat gct aag agt ggc aat gtt acc gcc cca act 2300
Thr Glu Ile Ser Val Asp Ala Lys Ser Gly Asn Val Thr Ala Pro Thr
520 525 530

tac aac att ggc gtg aaa acc acc gag ctt aac agt gat ggc act agt 2348
Tyr Asn Ile Gly Val Lys Thr Thr Glu Leu Asn Ser Asp Gly Thr Ser
535 540 545

gat aaa ttt agt gtt aag ggt agt ggt acg aac aat agc tta gtt acc 2396
Asp Lys Phe Ser Val Lys Gly Ser Gly Thr Asn Asn Ser Leu Val Thr
550 555 560

gcc gaa cat ttg gca agc tat cta aat gaa gtc aat cga acg gct gac 2444
Ala Glu His Leu Ala Ser Tyr Leu Asn Glu Val Asn Arg Thr Ala Asp
565 570 575

agt gct cta caa agc ttt acc gtt aaa gaa gaa gac gat gat gac gcc 2492
Ser Ala Leu Gln Ser Phe Thr Val Lys Glu Glu Asp Asp Asp Asp Ala
580 585 590 595

aac gct atc acc gtg gct aaa gat acg aca aaa aat gcc ggc gca gtc 2540

```
Asn Ala Ile Thr Val Ala Lys Asp Thr Thr Lys Asn Ala Gly Ala Val
                600                     605                 610

agc atc tta aaa ctc aaa ggt aaa aac ggt cta acg gtt gct acc aaa     2588
Ser Ile Leu Lys Leu Lys Gly Lys Asn Gly Leu Thr Val Ala Thr Lys
            615                 620                 625

aaa gat ggt acg gtt acc ttt ggg ctt agc caa gat agc ggt ctg acc     2636
Lys Asp Gly Thr Val Thr Phe Gly Leu Ser Gln Asp Ser Gly Leu Thr
        630                 635                 640

att ggc aaa agc acc cta aac aac gat ggc ttg act gtt aaa gat acc     2684
Ile Gly Lys Ser Thr Leu Asn Asn Asp Gly Leu Thr Val Lys Asp Thr
    645                 650                 655

aac gaa caa atc caa gtc ggt gct aat ggc att aaa ttt act aat gtg     2732
Asn Glu Gln Ile Gln Val Gly Ala Asn Gly Ile Lys Phe Thr Asn Val
660                 665                 670                 675

aat ggt agt aat cca ggt act ggc att gca aat acc gct cgc att acc     2780
Asn Gly Ser Asn Pro Gly Thr Gly Ile Ala Asn Thr Ala Arg Ile Thr
                680                 685                 690

aga gat aaa att ggc ttt gct ggt tct gat ggt gca gtt gat aca aac     2828
Arg Asp Lys Ile Gly Phe Ala Gly Ser Asp Gly Ala Val Asp Thr Asn
            695                 700                 705

aaa cct tat ctt gat caa gac aag cta caa gtt ggc aat gtt aag att     2876
Lys Pro Tyr Leu Asp Gln Asp Lys Leu Gln Val Gly Asn Val Lys Ile
        710                 715                 720

acc aac act ggc att aac gca ggt ggt aaa gcc atc aca ggg ctg tcc     2924
Thr Asn Thr Gly Ile Asn Ala Gly Gly Lys Ala Ile Thr Gly Leu Ser
    725                 730                 735

cca aca ctg cct agc att gcc gat caa agt agc cgc aac ata gaa ctg     2972
Pro Thr Leu Pro Ser Ile Ala Asp Gln Ser Ser Arg Asn Ile Glu Leu
740                 745                 750                 755

ggc aat aca atc caa gac aaa gac aaa tcc aac gct gcc agc att aat     3020
Gly Asn Thr Ile Gln Asp Lys Asp Lys Ser Asn Ala Ala Ser Ile Asn
                760                 765                 770

gat ata tta aat aca ggc ttt aac cta aaa aat aat aac aac ccc att     3068
Asp Ile Leu Asn Thr Gly Phe Asn Leu Lys Asn Asn Asn Asn Pro Ile
            775                 780                 785

gac ttt gtc tcc act tat gac att gtt gac ttt gcc aat ggc aat gcc     3116
Asp Phe Val Ser Thr Tyr Asp Ile Val Asp Phe Ala Asn Gly Asn Ala
        790                 795                 800

acc acc gcc aca gta acc cat gat acc gct aac aaa acc agt aaa gtg     3164
Thr Thr Ala Thr Val Thr His Asp Thr Ala Asn Lys Thr Ser Lys Val
    805                 810                 815

gta tat gat gtg aat gtg gat gat aca acc att cat cta aca ggc act     3212
Val Tyr Asp Val Asn Val Asp Asp Thr Thr Ile His Leu Thr Gly Thr
820                 825                 830                 835
```

gat gac aat aaa aaa ctt ggc gtc aaa acc acc aaa ctg aac aaa aca 3260
Asp Asp Asn Lys Lys Leu Gly Val Lys Thr Thr Lys Leu Asn Lys Thr
840 845 850

agt gct aat ggt aat aca gca act aac ttt aat gtt aac tct agt gat 3308
Ser Ala Asn Gly Asn Thr Ala Thr Asn Phe Asn Val Asn Ser Ser Asp
855 860 865

gaa gat gcc ctt gtt aac gcc aaa gac atc gcc gaa aat cta aac acc 3356
Glu Asp Ala Leu Val Asn Ala Lys Asp Ile Ala Glu Asn Leu Asn Thr
870 875 880

cta gcc aag gaa att cac acc acc aaa ggc aca gca gac acc gcc cta 3404
Leu Ala Lys Glu Ile His Thr Thr Lys Gly Thr Ala Asp Thr Ala Leu
885 890 895

caa acc ttt acc gtt aaa aag gta gat gaa aat aat aat gct gat gac 3452
Gln Thr Phe Thr Val Lys Lys Val Asp Glu Asn Asn Asn Ala Asp Asp
900 905 910 915

gcc aac gcc atc acc gtg ggt caa aag aac gca aat aat caa gtc aac 3500
Ala Asn Ala Ile Thr Val Gly Gln Lys Asn Ala Asn Asn Gln Val Asn
920 925 930

acc cta aca ctc aaa ggt gaa aac ggt ctt aat att aaa acc gac aaa 3548
Thr Leu Thr Leu Lys Gly Glu Asn Gly Leu Asn Ile Lys Thr Asp Lys
935 940 945

aat ggt acg gtt acc ttt ggc att aac acc aca agc ggt ctt aaa gcc 3596
Asn Gly Thr Val Thr Phe Gly Ile Asn Thr Thr Ser Gly Leu Lys Ala
950 955 960

ggc aaa agc acc cta aac gac ggt ggc ttg tct att aaa aac ccc act 3644
Gly Lys Ser Thr Leu Asn Asp Gly Gly Leu Ser Ile Lys Asn Pro Thr
965 970 975

ggt agc gaa caa atc caa gtc ggt gct gat ggc gtg aag ttt gcc aag 3692
Gly Ser Glu Gln Ile Gln Val Gly Ala Asp Gly Val Lys Phe Ala Lys
980 985 990 995

gtt aat aat aat ggt gtt gta ggt gct ggc att gat ggc aca act cgc 3740
Val Asn Asn Asn Gly Val Val Gly Ala Gly Ile Asp Gly Thr Thr Arg
1000 1005 1010

att acc aga gat gaa att ggc ttt act ggg act aat ggc tca ctt gat 3788
Ile Thr Arg Asp Glu Ile Gly Phe Thr Gly Thr Asn Gly Ser Leu Asp
1015 1020 1025

aaa agc aaa ccc cac cta agc aaa gac ggc att aac gca ggt ggt aaa 3836
Lys Ser Lys Pro His Leu Ser Lys Asp Gly Ile Asn Ala Gly Gly Lys
1030 1035 1040

aag att acc aac att caa tca ggt gag att gcc caa aac agc cat gat 3884
Lys Ile Thr Asn Ile Gln Ser Gly Glu Ile Ala Gln Asn Ser His Asp
1045 1050 1055

gct gtg aca ggc ggc aag att tat gat tta aaa acc gaa ctt gaa aac 3932
Ala Val Thr Gly Gly Lys Ile Tyr Asp Leu Lys Thr Glu Leu Glu Asn
1060 1065 1070 1075

aaa atc agc agt act gcc aaa aca gca caa aac tca tta cac gaa ttc 3980
Lys Ile Ser Ser Thr Ala Lys Thr Ala Gln Asn Ser Leu His Glu Phe
1080 1085 1090

tca gta gca gat gaa caa ggt aat aac ttt acg gtt agt aac cct tac 4028
Ser Val Ala Asp Glu Gln Gly Asn Asn Phe Thr Val Ser Asn Pro Tyr
1095 1100 1105

tcc agt tat gac acc tca aag acc tct gat gtc atc acc ttt gca ggt 4076
Ser Ser Tyr Asp Thr Ser Lys Thr Ser Asp Val Ile Thr Phe Ala Gly
1110 1115 1120

gaa aac ggc att acc acc aag gta aat aaa ggt gtg gtg cgt gtg ggc 4124
Glu Asn Gly Ile Thr Thr Lys Val Asn Lys Gly Val Val Arg Val Gly
1125 1130 1135

att gac caa acc aaa ggc tta acc acg cct aag ctg acc gtg ggt aat 4172
Ile Asp Gln Thr Lys Gly Leu Thr Thr Pro Lys Leu Thr Val Gly Asn
1140 1145 1150 1155

aat aat ggc aaa ggc att gtc att gac agc caa aat ggt caa aat acc 4220
Asn Asn Gly Lys Gly Ile Val Ile Asp Ser Gln Asn Gly Gln Asn Thr
1160 1165 1170

atc aca gga cta agc aac act cta gct aat gtt acc aat gat aaa ggt 4268
Ile Thr Gly Leu Ser Asn Thr Leu Ala Asn Val Thr Asn Asp Lys Gly
1175 1180 1185

agc gta cgc acc aca gaa cag ggc aat ata atc aaa gac gaa gac aaa 4316
Ser Val Arg Thr Thr Glu Gln Gly Asn Ile Ile Lys Asp Glu Asp Lys
1190 1195 1200

acc cgt gcc gcc agc att gtt gat gtg cta agc gca ggc ttt aac ttg 4364
Thr Arg Ala Ala Ser Ile Val Asp Val Leu Ser Ala Gly Phe Asn Leu
1205 1210 1215

caa ggc aat ggt gaa gcg gtt gac ttt gtc tcc act tat gac acc gtc 4412
Gln Gly Asn Gly Glu Ala Val Asp Phe Val Ser Thr Tyr Asp Thr Val
1220 1225 1230 1235

aac ttt gcc gat ggc aat gcc acc acc gct aag gtg acc tat gat gac 4460
Asn Phe Ala Asp Gly Asn Ala Thr Thr Ala Lys Val Thr Tyr Asp Asp
1240 1245 1250

aca agc aaa acc agt aaa gtg gtc tat gat gtc aat gtg gat gat aca 4508
Thr Ser Lys Thr Ser Lys Val Val Tyr Asp Val Asn Val Asp Asp Thr
1255 1260 1265

acc att gaa gtt aaa gat aaa aaa ctt ggc gta aaa acc acc aca ttg 4556
Thr Ile Glu Val Lys Asp Lys Lys Leu Gly Val Lys Thr Thr Thr Leu
1270 1275 1280

acc agt act ggc aca ggt gct aat aaa ttt gcc cta agc aat caa gct 4604
Thr Ser Thr Gly Thr Gly Ala Asn Lys Phe Ala Leu Ser Asn Gln Ala
1285 1290 1295

act ggc gat gcg ctt gtc aag gcc agt gat atc gtt gct cat cta aac 4652
Thr Gly Asp Ala Leu Val Lys Ala Ser Asp Ile Val Ala His Leu Asn

1300 1305 1310 1315

acc tta tct ggc gac atc caa act gcc aaa ggg gca agc caa gcg aac 4700
Thr Leu Ser Gly Asp Ile Gln Thr Ala Lys Gly Ala Ser Gln Ala Asn
1320 1325 1330

aac tca gca ggc tat gtg gat gct gat ggc aat aag gtc atc tat gac 4748
Asn Ser Ala Gly Tyr Val Asp Ala Asp Gly Asn Lys Val Ile Tyr Asp
1335 1340 1345

agt acc gat aac aag tac tat caa gcc aaa aat gat ggc aca gtt gat 4796
Ser Thr Asp Asn Lys Tyr Tyr Gln Ala Lys Asn Asp Gly Thr Val Asp
1350 1355 1360

aaa acc aaa gaa gtt gcc aaa gac aaa ctg gtc gcc caa gcc caa acc 4844
Lys Thr Lys Glu Val Ala Lys Asp Lys Leu Val Ala Gln Ala Gln Thr
1365 1370 1375

cca gat ggc aca ttg gct caa atg aat gtc aaa tca gtc att aac aaa 4892
Pro Asp Gly Thr Leu Ala Gln Met Asn Val Lys Ser Val Ile Asn Lys
1380 1385 1390 1395

gaa caa gta aat gat gcc aat aaa aag caa ggc atc aat gaa gac aac 4940
Glu Gln Val Asn Asp Ala Asn Lys Lys Gln Gly Ile Asn Glu Asp Asn
1400 1405 1410

gcc ttt gtt aaa gga ctt gaa aaa gcc gct tct gat aac aaa acc aaa 4988
Ala Phe Val Lys Gly Leu Glu Lys Ala Ala Ser Asp Asn Lys Thr Lys
1415 1420 1425

aac gcc gca gta act gtg ggt gat tta aat gcc gtt gcc caa aca ccg 5036
Asn Ala Ala Val Thr Val Gly Asp Leu Asn Ala Val Ala Gln Thr Pro
1430 1435 1440

ctg acc ttt gca ggg gat aca ggc aca acg gct aaa aaa ctg ggc gag 5084
Leu Thr Phe Ala Gly Asp Thr Gly Thr Thr Ala Lys Lys Leu Gly Glu
1445 1450 1455

act ttg acc atc aaa ggt ggg caa aca gac acc aat aag cta acc gat 5132
Thr Leu Thr Ile Lys Gly Gly Gln Thr Asp Thr Asn Lys Leu Thr Asp
1460 1465 1470 1475

aat aac atc ggt gtg gta gca ggt act gat ggc ttc act gtc aaa ctt 5180
Asn Asn Ile Gly Val Val Ala Gly Thr Asp Gly Phe Thr Val Lys Leu
1480 1485 1490

gcc aaa gac cta acc aat ctt aac agc gtt aat gca ggt ggc acc aaa 5228
Ala Lys Asp Leu Thr Asn Leu Asn Ser Val Asn Ala Gly Gly Thr Lys
1495 1500 1505

att gat gac aaa ggc gtg tct ttt gta gac tca agc ggt caa gcc aaa 5276
Ile Asp Asp Lys Gly Val Ser Phe Val Asp Ser Ser Gly Gln Ala Lys
1510 1515 1520

gca aac acc cct gtg cta agt gcc aat ggg ctg gac ctg ggt ggc aag 5324
Ala Asn Thr Pro Val Leu Ser Ala Asn Gly Leu Asp Leu Gly Gly Lys
1525 1530 1535

gtc atc agt aat gtg ggc aaa ggc aca aaa gat acc gac gct gcc aat 5372

Val Ile Ser Asn Val Gly Lys Gly Thr Lys Asp Thr Asp Ala Ala Asn
1540 1545 1550 1555

gta caa cag tta aac gaa gta cgc aac ttg ttg ggt ctt ggt aat gct 5420
Val Gln Gln Leu Asn Glu Val Arg Asn Leu Leu Gly Leu Gly Asn Ala
1560 1565 1570

ggt aat gat aac gct gac ggc aat cag gta aac att gcc gac atc aaa 5468
Gly Asn Asp Asn Ala Asp Gly Asn Gln Val Asn Ile Ala Asp Ile Lys
1575 1580 1585

aaa gac cca aat tca ggt tca tca tct aac cgc act gtc atc aaa gca 5516
Lys Asp Pro Asn Ser Gly Ser Ser Ser Asn Arg Thr Val Ile Lys Ala
1590 1595 1600

ggc acg gta ctt ggc ggt aaa ggt aat aac gat acc gaa aaa ctt gcc 5564
Gly Thr Val Leu Gly Gly Lys Gly Asn Asn Asp Thr Glu Lys Leu Ala
1605 1610 1615

act ggt ggt ata caa gtg ggc gtg gat aaa gac ggc aac gct aac ggc 5612
Thr Gly Gly Ile Gln Val Gly Val Asp Lys Asp Gly Asn Ala Asn Gly
1620 1625 1630 1635

gat tta agc aat gtt tgg gtc aaa acc caa aaa gat ggc agc aaa aaa 5660
Asp Leu Ser Asn Val Trp Val Lys Thr Gln Lys Asp Gly Ser Lys Lys
1640 1645 1650

gcc ctg ctc gcc act tat aac gcc gca ggt cag acc aac tat ttg acc 5708
Ala Leu Leu Ala Thr Tyr Asn Ala Ala Gly Gln Thr Asn Tyr Leu Thr
1655 1660 1665

aac aac ccc gca gaa gcc att gac aga ata aat gaa caa ggt atc cgc 5756
Asn Asn Pro Ala Glu Ala Ile Asp Arg Ile Asn Glu Gln Gly Ile Arg
1670 1675 1680

ttc ttc cat gtc aac gat ggc aat caa gag cct gtg gta caa ggg cgt 5804
Phe Phe His Val Asn Asp Gly Asn Gln Glu Pro Val Val Gln Gly Arg
1685 1690 1695

aac ggc att gac tca agt gcc tca ggc aag cac tca gtg gcg ata ggt 5852
Asn Gly Ile Asp Ser Ser Ala Ser Gly Lys His Ser Val Ala Ile Gly
1700 1705 1710 1715

ttc cag gcc aag gca gat ggt gaa gcc gcc gtt gcc ata ggc aga caa 5900
Phe Gln Ala Lys Ala Asp Gly Glu Ala Ala Val Ala Ile Gly Arg Gln
1720 1725 1730

acc caa gca ggc aac caa tcc atc gcc atc ggt gat aac gca caa gcc 5948
Thr Gln Ala Gly Asn Gln Ser Ile Ala Ile Gly Asp Asn Ala Gln Ala
1735 1740 1745

acg ggc gat caa tcc atc gcc atc ggt aca ggc aat gtg gta gca ggt 5996
Thr Gly Asp Gln Ser Ile Ala Ile Gly Thr Gly Asn Val Val Ala Gly
1750 1755 1760

aag cac tct ggt gcc atc ggc gac cca agc act gtt aag gct gat aac 6044
Lys His Ser Gly Ala Ile Gly Asp Pro Ser Thr Val Lys Ala Asp Asn
1765 1770 1775

agt tac agt gtg ggt aat aac aac cag ttt acc gat gcc act caa acc 6092
Ser Tyr Ser Val Gly Asn Asn Asn Gln Phe Thr Asp Ala Thr Gln Thr
1780 1785 1790 1795

gat gtc ttt ggt gtg ggc aat aac atc acc gtg acc gaa agt aac tcg 6140
Asp Val Phe Gly Val Gly Asn Asn Ile Thr Val Thr Glu Ser Asn Ser
1800 1805 1810

gtt gcc tta ggt tca aac tct gcc atc agt gca ggc aca cac gca ggc 6188
Val Ala Leu Gly Ser Asn Ser Ala Ile Ser Ala Gly Thr His Ala Gly
1815 1820 1825

aca caa gcc aaa aaa tct gac ggc aca gca ggt aca acc acc aca gca 6236
Thr Gln Ala Lys Lys Ser Asp Gly Thr Ala Gly Thr Thr Thr Thr Ala
1830 1835 1840

ggt gca acc ggt acg gtt aaa ggc ttt gct gga caa acg gcg gtt ggt 6284
Gly Ala Thr Gly Thr Val Lys Gly Phe Ala Gly Gln Thr Ala Val Gly
1845 1850 1855

gcg gtc tcc gtg ggt gcc tca ggt gct gaa cgc cgt atc caa aat gtg 6332
Ala Val Ser Val Gly Ala Ser Gly Ala Glu Arg Arg Ile Gln Asn Val
1860 1865 1870 1875

gca gca ggt gag gtc agt gcc acc agc acc gat gcg gtc aat ggt agc 6380
Ala Ala Gly Glu Val Ser Ala Thr Ser Thr Asp Ala Val Asn Gly Ser
1880 1885 1890

cag ttg tac aaa gcc acc caa agc att gcc aac gca acc aat gag ctt 6428
Gln Leu Tyr Lys Ala Thr Gln Ser Ile Ala Asn Ala Thr Asn Glu Leu
1895 1900 1905

gac cat cgt atc cac caa aac gaa aat aag gcc aat gca ggg att tca 6476
Asp His Arg Ile His Gln Asn Glu Asn Lys Ala Asn Ala Gly Ile Ser
1910 1915 1920

tca gcg atg gcg atg gcg tcc atg cca caa gcc tac att cct ggc aga 6524
Ser Ala Met Ala Met Ala Ser Met Pro Gln Ala Tyr Ile Pro Gly Arg
1925 1930 1935

tcc atg gtt acc ggg ggt att gcc acc cac aac ggt caa ggt gcg gtg 6572
Ser Met Val Thr Gly Gly Ile Ala Thr His Asn Gly Gln Gly Ala Val
1940 1945 1950 1955

gca gtg gga ctg tcg aag ctg tcg gat aat ggt caa tgg gta ttt aaa 6620
Ala Val Gly Leu Ser Lys Leu Ser Asp Asn Gly Gln Trp Val Phe Lys
1960 1965 1970

atc aat ggt tca gcc gat acc caa ggc cat gta ggg gcg gca gtt ggt 6668
Ile Asn Gly Ser Ala Asp Thr Gln Gly His Val Gly Ala Ala Val Gly
1975 1980 1985

gca ggt ttt cac ttt taagccataa atcgcaagat tttacttaaa aatcaatctc 6723
Ala Gly Phe His Phe
1990

accatagttg tataaaacag catcagcatc agtcatatta ctgatgctga tgtttttat 6783

cacttaaacc attttaccgc tcaagtgatt ctctttcacc atgaccaaat cgccattgat 6843

cataggtaaa cttattgagt aaattttatc aatgtagttg ttagatatgg ttaaaattgt 6903

gccattgacc aaaaaatgac cgatttatcc cgaaaatttc tgattatgat ccgttgacct 6963

gcaggtcgac 6973

Figure 3. *M. catarrhalis* strain 4223 genomic 200kDa gene.

```
ccatggatat gggcaggtgt gctcgcctgc cgtatgatgg cgatgacacc ccatttgccc 60

catatctgta cgatttgaca tgtgatatga tttaacatgt gacatgattt aacattgttt 120

aatactgttg ccatcattac cataatttag taacgcattt agtaacgcat ttgtaaaaat 180

cattgcgccc ctttatgtgt atcatatgaa tagaatatta tgattgtatc tgattattgt 240

atcagaatgg tgatgctata tgatgatgcc tacgagttga tttgggttaa tcactctatg 300

atttgatata ttttgaaact aatctattga cttaaatcac catatggtta taatttagca 360

taatggtagg ctttttgtaa aaatcacatc gcaatattgt tctactgtta ctaccatgct 420

tgaatgacga tcccaatcac cagattcatt caagtgatgt gtttgtatac gcaccattta 480

ccctaattat ttcaatcaaa tgcctatgtc agcatgtatc atttttttaa ggtaaaccac 540

c atg aat cac atc tat aaa gtc atc ttt aac aaa gcc aca ggc aca ttt 589
  Met Asn His Ile Tyr Lys Val Ile Phe Asn Lys Ala Thr Gly Thr Phe
    1               5                  10                  15

atg gca gtg gca gag tac gcc aaa tcc cac agc acg ggg ggg ggt agc   637
Met Ala Val Ala Glu Tyr Ala Lys Ser His Ser Thr Gly Gly Gly Ser
             20                  25                  30

tgt gct aca ggg caa gtt ggc agt gta tgc act ctg agc ttt gcc cgt   685
Cys Ala Thr Gly Gln Val Gly Ser Val Cys Thr Leu Ser Phe Ala Arg
         35                  40                  45

att gcc gcg ctc gct gtc ctc gtg atc ggt gca acg ctc agt ggc agt   733
Ile Ala Ala Leu Ala Val Leu Val Ile Gly Ala Thr Leu Ser Gly Ser
     50                  55                  60

gct tat gct caa aaa aaa gat acc aaa cat atc gca att ggt gaa caa   781
Ala Tyr Ala Gln Lys Lys Asp Thr Lys His Ile Ala Ile Gly Glu Gln
 65                  70                  75                  80

aac cag cca aga cgc tca ggc act gcc aag gcg gac ggt gat cga gcc   829
Asn Gln Pro Arg Arg Ser Gly Thr Ala Lys Ala Asp Gly Asp Arg Ala
                 85                  90                  95

att gct att ggt gaa aat gct aac gca cag ggc ggt caa gcc atc gcc   877
Ile Ala Ile Gly Glu Asn Ala Asn Ala Gln Gly Gly Gln Ala Ile Ala
            100                 105                 110

atc ggt agt agt aat aaa act gtc aat gga agc agt ttg gat aag ata   925
Ile Gly Ser Ser Asn Lys Thr Val Asn Gly Ser Ser Leu Asp Lys Ile
        115                 120                 125

ggt acc gat gct acg ggt caa gag tcc atc gcc atc ggt ggt gat gta   973
Gly Thr Asp Ala Thr Gly Gln Glu Ser Ile Ala Ile Gly Gly Asp Val
    130                 135                 140

aag gct agt ggt gat gcc tcg att gcc atc ggt agt gat gac tta cat   1021
Lys Ala Ser Gly Asp Ala Ser Ile Ala Ile Gly Ser Asp Asp Leu His
```

```
145                 150                 155                 160

ttg ctt gat cag cat ggt aat cct aaa cat ccg aaa ggt act ctg att     1069
Leu Leu Asp Gln His Gly Asn Pro Lys His Pro Lys Gly Thr Leu Ile
                165                 170                 175

aac gat ctt att aac ggc cat gca gta tta aaa gaa ata cga agc tca     1117
Asn Asp Leu Ile Asn Gly His Ala Val Leu Lys Glu Ile Arg Ser Ser
            180                 185                 190

aag gat aat gat gta aaa tat aga cgc aca acc gca agc gga cac gcc     1165
Lys Asp Asn Asp Val Lys Tyr Arg Arg Thr Thr Ala Ser Gly His Ala
        195                 200                 205

agt act gca gtg gga gcc atg tca tat gca cag ggt cat ttt tcc aac     1213
Ser Thr Ala Val Gly Ala Met Ser Tyr Ala Gln Gly His Phe Ser Asn
    210                 215                 220

gcc ttt ggt aca cgg gca aca gct aaa agt gcc tat tcc ttg gca gtg     1261
Ala Phe Gly Thr Arg Ala Thr Ala Lys Ser Ala Tyr Ser Leu Ala Val
225                 230                 235                 240

ggt ctt gcc gcc aca gcc gag ggc caa tct aca atc gct att ggt tct     1309
Gly Leu Ala Ala Thr Ala Glu Gly Gln Ser Thr Ile Ala Ile Gly Ser
                245                 250                 255

gat gca aca tct agc tcg ttg gga gcg ata gcc ctt ggt gca ggt act     1357
Asp Ala Thr Ser Ser Ser Leu Gly Ala Ile Ala Leu Gly Ala Gly Thr
            260                 265                 270

cgt gct cag cta cag ggc agt att gcc cta ggt caa ggt tct gtt gtc     1405
Arg Ala Gln Leu Gln Gly Ser Ile Ala Leu Gly Gln Gly Ser Val Val
        275                 280                 285

act cag agt gat aat aat tct aga ccg gcc tat aca cca aat acc cag     1453
Thr Gln Ser Asp Asn Asn Ser Arg Pro Ala Tyr Thr Pro Asn Thr Gln
    290                 295                 300

gca cta gac ccc aag ttt caa gcc acc aat aat acg aag gcg ggt cca     1501
Ala Leu Asp Pro Lys Phe Gln Ala Thr Asn Asn Thr Lys Ala Gly Pro
305                 310                 315                 320

ctt tcc att ggt agt aac tct atc aaa cgt aaa atc atc aat gtc ggt     1549
Leu Ser Ile Gly Ser Asn Ser Ile Lys Arg Lys Ile Ile Asn Val Gly
                325                 330                 335

gca ggt gtt aat aaa acc gat gcg gtc aat gtg gca cag cta gaa gcg     1597
Ala Gly Val Asn Lys Thr Asp Ala Val Asn Val Ala Gln Leu Glu Ala
            340                 345                 350

gtg gtg aag tgg gct aag gag cgt aga att act ttt cag ggt gat gat     1645
Val Val Lys Trp Ala Lys Glu Arg Arg Ile Thr Phe Gln Gly Asp Asp
        355                 360                 365

aac agt act gac gta aaa ata ggt ttg gat aat act tta act att aaa     1693
Asn Ser Thr Asp Val Lys Ile Gly Leu Asp Asn Thr Leu Thr Ile Lys
    370                 375                 380

ggt ggt gca gag acc aac gca tta acc gat aat aat atc ggt gtg gta     1741
```

Gly Gly Ala Glu Thr Asn Ala Leu Thr Asp Asn Asn Ile Gly Val Val
385 390 395 400

aaa gag gct gat aat agt ggt ctg aaa gtt aaa ctt gct aaa act tta 1789
Lys Glu Ala Asp Asn Ser Gly Leu Lys Val Lys Leu Ala Lys Thr Leu
405 410 415

aac aat ctt act gag gtg aat aca act aca tta aat gcc aca acc aca 1837
Asn Asn Leu Thr Glu Val Asn Thr Thr Thr Leu Asn Ala Thr Thr Thr
420 425 430

gtt aag gta ggt agt agt agt agt act aca gct gaa tta ttg agt gat 1885
Val Lys Val Gly Ser Ser Ser Ser Thr Thr Ala Glu Leu Leu Ser Asp
435 440 445

agt tta acc ttt acc cag ccc aat aca ggc agt caa agc aca agc aaa 1933
Ser Leu Thr Phe Thr Gln Pro Asn Thr Gly Ser Gln Ser Thr Ser Lys
450 455 460

acc gtc tat ggc gtt aat ggg gtg aag ttt act aat aat gca gaa aca 1981
Thr Val Tyr Gly Val Asn Gly Val Lys Phe Thr Asn Asn Ala Glu Thr
465 470 475 480

aca gca gca atc ggc act act cgt att acc aga gat aaa att ggc ttt 2029
Thr Ala Ala Ile Gly Thr Thr Arg Ile Thr Arg Asp Lys Ile Gly Phe
485 490 495

gct cga gat ggt gat gtt gat gaa aaa caa gca cca tat ttg gat aaa 2077
Ala Arg Asp Gly Asp Val Asp Glu Lys Gln Ala Pro Tyr Leu Asp Lys
500 505 510

aaa caa ctt aaa gtg ggt agt gtt gca att acc ata gac aat ggc att 2125
Lys Gln Leu Lys Val Gly Ser Val Ala Ile Thr Ile Asp Asn Gly Ile
515 520 525

gat gca ggt aat aaa aag atc agt aat ctt gcc aaa ggt agc agt gct 2173
Asp Ala Gly Asn Lys Lys Ile Ser Asn Leu Ala Lys Gly Ser Ser Ala
530 535 540

aac gat gcg gtt acc atc gaa cag ctc aaa gcc gcc aag cct act tta 2221
Asn Asp Ala Val Thr Ile Glu Gln Leu Lys Ala Ala Lys Pro Thr Leu
545 550 555 560

aac gca ggc gct ggc atc agt gtc aca cct act gaa ata tca gtt gat 2269
Asn Ala Gly Ala Gly Ile Ser Val Thr Pro Thr Glu Ile Ser Val Asp
565 570 575

gct aag agt ggc aat gtt acc gcc cca act tac aac att ggc gtg aaa 2317
Ala Lys Ser Gly Asn Val Thr Ala Pro Thr Tyr Asn Ile Gly Val Lys
580 585 590

acc acc gag ctt aac agt gat ggc act agt gat aaa ttt agt gtt aag 2365
Thr Thr Glu Leu Asn Ser Asp Gly Thr Ser Asp Lys Phe Ser Val Lys
595 600 605

ggt agt ggt acg aac aat agc tta gtt acc gcc gaa cat ttg gca agc 2413
Gly Ser Gly Thr Asn Asn Ser Leu Val Thr Ala Glu His Leu Ala Ser
610 615 620

tat cta aat gaa gtc aat cga acg gct gac agt gct cta caa agc ttt 2461
Tyr Leu Asn Glu Val Asn Arg Thr Ala Asp Ser Ala Leu Gln Ser Phe
625 630 635 640

acc gtt aaa gaa gaa gac gat gat gac gcc aac gct atc acc gtg gct 2509
Thr Val Lys Glu Glu Asp Asp Asp Asp Ala Asn Ala Ile Thr Val Ala
645 650 655

aaa gat acg aca aaa aat gcc ggc gca gtc agc atc tta aaa ctc aaa 2557
Lys Asp Thr Thr Lys Asn Ala Gly Ala Val Ser Ile Leu Lys Leu Lys
660 665 670

ggt aaa aac ggt cta acg gtt gct acc aaa aaa gat ggt acg gtt acc 2605
Gly Lys Asn Gly Leu Thr Val Ala Thr Lys Lys Asp Gly Thr Val Thr
675 680 685

ttt ggg ctt agc caa gat agc ggt ctg acc att ggc aaa agc acc cta 2653
Phe Gly Leu Ser Gln Asp Ser Gly Leu Thr Ile Gly Lys Ser Thr Leu
690 695 700

aac aac gat ggc ttg act gtt aaa gat acc aac gaa caa atc caa gtc 2701
Asn Asn Asp Gly Leu Thr Val Lys Asp Thr Asn Glu Gln Ile Gln Val
705 710 715 720

ggt gct aat ggc att aaa ttt act aat gtg aat ggt agt aat cca ggt 2749
Gly Ala Asn Gly Ile Lys Phe Thr Asn Val Asn Gly Ser Asn Pro Gly
725 730 735

act ggc att gca aat acc gct cgc att acc aga gat aaa att ggc ttt 2797
Thr Gly Ile Ala Asn Thr Ala Arg Ile Thr Arg Asp Lys Ile Gly Phe
740 745 750

gct ggt tct gat ggt gca gtt gat aca aac aaa cct tat ctt gat caa 2845
Ala Gly Ser Asp Gly Ala Val Asp Thr Asn Lys Pro Tyr Leu Asp Gln
755 760 765

gac aag cta caa gtt ggc aat gtt aag att acc aac act ggc att aac 2893
Asp Lys Leu Gln Val Gly Asn Val Lys Ile Thr Asn Thr Gly Ile Asn
770 775 780

gca ggt ggt aaa gcc atc aca ggg ctg tcc cca aca ctg cct agc att 2941
Ala Gly Gly Lys Ala Ile Thr Gly Leu Ser Pro Thr Leu Pro Ser Ile
785 790 795 800

gcc gat caa agt agc cgc aac ata gaa ctg ggc aat aca atc caa gac 2989
Ala Asp Gln Ser Ser Arg Asn Ile Glu Leu Gly Asn Thr Ile Gln Asp
805 810 815

aaa gac aaa tcc aac gct gcc agc att aat gat ata tta aat aca ggc 3037
Lys Asp Lys Ser Asn Ala Ala Ser Ile Asn Asp Ile Leu Asn Thr Gly
820 825 830

ttt aac cta aaa aat aat aac aac ccc att gac ttt gtc tcc act tat 3085
Phe Asn Leu Lys Asn Asn Asn Asn Pro Ile Asp Phe Val Ser Thr Tyr
835 840 845

gac att gtt gac ttt gcc aat ggc aat gcc acc acc gcc aca gta acc 3133
Asp Ile Val Asp Phe Ala Asn Gly Asn Ala Thr Thr Ala Thr Val Thr
850 855 860

cat gat acc gct aac aaa acc agt aaa gtg gta tat gat gtg aat gtg 3181
His Asp Thr Ala Asn Lys Thr Ser Lys Val Val Tyr Asp Val Asn Val
865 870 875 880

gat gat aca acc att cat cta aca ggc act gat gac aat aaa aaa ctt 3229
Asp Asp Thr Thr Ile His Leu Thr Gly Thr Asp Asp Asn Lys Lys Leu
885 890 895

ggc gtc aaa acc acc aaa ctg aac aaa aca agt gct aat ggt aat aca 3277
Gly Val Lys Thr Thr Lys Leu Asn Lys Thr Ser Ala Asn Gly Asn Thr
900 905 910

gca act aac ttt aat gtt aac tct agt gat gaa gat gcc ctt gtt aac 3325
Ala Thr Asn Phe Asn Val Asn Ser Ser Asp Glu Asp Ala Leu Val Asn
915 920 925

gcc aaa gac atc gcc gaa aat cta aac acc cta gcc aag gaa att cac 3373
Ala Lys Asp Ile Ala Glu Asn Leu Asn Thr Leu Ala Lys Glu Ile His
930 935 940

acc acc aaa ggc aca gca gac acc gcc cta caa acc ttt acc gtt aaa 3421
Thr Thr Lys Gly Thr Ala Asp Thr Ala Leu Gln Thr Phe Thr Val Lys
945 950 955 960

aag gta gat gaa aat aat aat gct gat gac gcc aac gcc atc acc gtg 3469
Lys Val Asp Glu Asn Asn Asn Ala Asp Asp Ala Asn Ala Ile Thr Val
965 970 975

ggt caa aag aac gca aat aat caa gtc aac acc cta aca ctc aaa ggt 3517
Gly Gln Lys Asn Ala Asn Asn Gln Val Asn Thr Leu Thr Leu Lys Gly
980 985 990

gaa aac ggt ctt aat att aaa acc gac aaa aat ggt acg gtt acc ttt 3565
Glu Asn Gly Leu Asn Ile Lys Thr Asp Lys Asn Gly Thr Val Thr Phe
995 1000 1005

ggc att aac acc aca agc ggt ctt aaa gcc ggc aaa agc acc cta aac 3613
Gly Ile Asn Thr Thr Ser Gly Leu Lys Ala Gly Lys Ser Thr Leu Asn
1010 1015 1020

gac ggt ggc ttg tct att aaa aac ccc act ggt agc gaa caa atc caa 3661
Asp Gly Gly Leu Ser Ile Lys Asn Pro Thr Gly Ser Glu Gln Ile Gln
1025 1030 1035 1040

gtc ggt gct gat ggc gtg aag ttt gcc aag gtt aat aat aat ggt gtt 3709
Val Gly Ala Asp Gly Val Lys Phe Ala Lys Val Asn Asn Asn Gly Val
1045 1050 1055

gta ggt gct ggc att gat ggc aca act cgc att acc aga gat gaa att 3757
Val Gly Ala Gly Ile Asp Gly Thr Thr Arg Ile Thr Arg Asp Glu Ile
1060 1065 1070

ggc ttt act ggg act aat ggc tca ctt gat aaa agc aaa ccc cac cta 3805
Gly Phe Thr Gly Thr Asn Gly Ser Leu Asp Lys Ser Lys Pro His Leu
1075 1080 1085

agc aaa gac ggc att aac gca ggt ggt aaa aag att acc aac att caa 3853
Ser Lys Asp Gly Ile Asn Ala Gly Gly Lys Lys Ile Thr Asn Ile Gln

1090 1095 1100

tca ggt gag att gcc caa aac agc cat gat gct gtg aca ggc ggc aag 3901
Ser Gly Glu Ile Ala Gln Asn Ser His Asp Ala Val Thr Gly Gly Lys
1105 1110 1115 1120

att tat gat tta aaa acc gaa ctt gaa aac aaa atc agc agt act gcc 3949
Ile Tyr Asp Leu Lys Thr Glu Leu Glu Asn Lys Ile Ser Ser Thr Ala
1125 1130 1135

aaa aca gca caa aac tca tta cac gaa ttc tca gta gca gat gaa caa 3997
Lys Thr Ala Gln Asn Ser Leu His Glu Phe Ser Val Ala Asp Glu Gln
1140 1145 1150

ggt aat aac ttt acg gtt agt aac cct tac tcc agt tat gac acc tca 4045
Gly Asn Asn Phe Thr Val Ser Asn Pro Tyr Ser Ser Tyr Asp Thr Ser
1155 1160 1165

aag acc tct gat gtc atc acc ttt gca ggt gaa aac ggc att acc acc 4093
Lys Thr Ser Asp Val Ile Thr Phe Ala Gly Glu Asn Gly Ile Thr Thr
1170 1175 1180

aag gta aat aaa ggt gtg gtg cgt gtg ggc att gac caa acc aaa ggc 4141
Lys Val Asn Lys Gly Val Val Arg Val Gly Ile Asp Gln Thr Lys Gly
1185 1190 1195 1200

tta acc acg cct aag ctg acc gtg ggt aat aat aat ggc aaa ggc att 4189
Leu Thr Thr Pro Lys Leu Thr Val Gly Asn Asn Asn Gly Lys Gly Ile
1205 1210 1215

gtc att gac agc caa aat ggt caa aat acc atc aca gga cta agc aac 4237
Val Ile Asp Ser Gln Asn Gly Gln Asn Thr Ile Thr Gly Leu Ser Asn
1220 1225 1230

act cta gct aat gtt acc aat gat aaa ggt agc gta cgc acc aca gaa 4285
Thr Leu Ala Asn Val Thr Asn Asp Lys Gly Ser Val Arg Thr Thr Glu
1235 1240 1245

cag ggc aat ata atc aaa gac gaa gac aaa acc cgt gcc gcc agc att 4333
Gln Gly Asn Ile Ile Lys Asp Glu Asp Lys Thr Arg Ala Ala Ser Ile
1250 1255 1260

gtt gat gtg cta agc gca ggc ttt aac ttg caa ggc aat ggt gaa gcg 4381
Val Asp Val Leu Ser Ala Gly Phe Asn Leu Gln Gly Asn Gly Glu Ala
1265 1270 1275 1280

gtt gac ttt gtc tcc act tat gac acc gtc aac ttt gcc gat ggc aat 4429
Val Asp Phe Val Ser Thr Tyr Asp Thr Val Asn Phe Ala Asp Gly Asn
1285 1290 1295

gcc acc acc gct aag gtg acc tat gat gac aca agc aaa acc agt aaa 4477
Ala Thr Thr Ala Lys Val Thr Tyr Asp Asp Thr Ser Lys Thr Ser Lys
1300 1305 1310

gtg gtc tat gat gtc aat gtg gat gat aca acc att gaa gtt aaa gat 4525
Val Val Tyr Asp Val Asn Val Asp Asp Thr Thr Ile Glu Val Lys Asp
1315 1320 1325

aaa aaa ctt ggc gta aaa acc acc aca ttg acc agt act ggc aca ggt 4573

Lys Lys Leu Gly Val Lys Thr Thr Thr Leu Thr Ser Thr Gly Thr Gly
1330 1335 1340

gct aat aaa ttt gcc cta agc aat caa gct act ggc gat gcg ctt gtc 4621
Ala Asn Lys Phe Ala Leu Ser Asn Gln Ala Thr Gly Asp Ala Leu Val
1345 1350 1355 1360

aag gcc agt gat atc gtt gct cat cta aac acc tta tct ggc gac atc 4669
Lys Ala Ser Asp Ile Val Ala His Leu Asn Thr Leu Ser Gly Asp Ile
1365 1370 1375

caa act gcc aaa ggg gca agc caa gcg aac aac tca gca ggc tat gtg 4717
Gln Thr Ala Lys Gly Ala Ser Gln Ala Asn Asn Ser Ala Gly Tyr Val
1380 1385 1390

gat gct gat ggc aat aag gtc atc tat gac agt acc gat aac aag tac 4765
Asp Ala Asp Gly Asn Lys Val Ile Tyr Asp Ser Thr Asp Asn Lys Tyr
1395 1400 1405

tat caa gcc aaa aat gat ggc aca gtt gat aaa acc aaa gaa gtt gcc 4813
Tyr Gln Ala Lys Asn Asp Gly Thr Val Asp Lys Thr Lys Glu Val Ala
1410 1415 1420

aaa gac aaa ctg gtc gcc caa gcc caa acc cca gat ggc aca ttg gct 4861
Lys Asp Lys Leu Val Ala Gln Ala Gln Thr Pro Asp Gly Thr Leu Ala
1425 1430 1435 1440

caa atg aat gtc aaa tca gtc att aac aaa gaa caa gta aat gat gcc 4909
Gln Met Asn Val Lys Ser Val Ile Asn Lys Glu Gln Val Asn Asp Ala
1445 1450 1455

aat aaa aag caa ggc atc aat gaa gac aac gcc ttt gtt aaa gga ctt 4957
Asn Lys Lys Gln Gly Ile Asn Glu Asp Asn Ala Phe Val Lys Gly Leu
1460 1465 1470

gaa aaa gcc gct tct gat aac aaa acc aaa aac gcc gca gta act gtg 5005
Glu Lys Ala Ala Ser Asp Asn Lys Thr Lys Asn Ala Ala Val Thr Val
1475 1480 1485

ggt gat tta aat gcc gtt gcc caa aca ccg ctg acc ttt gca ggg gat 5053
Gly Asp Leu Asn Ala Val Ala Gln Thr Pro Leu Thr Phe Ala Gly Asp
1490 1495 1500

aca ggc aca acg gct aaa aaa ctg ggc gag act ttg acc atc aaa ggt 5101
Thr Gly Thr Thr Ala Lys Lys Leu Gly Glu Thr Leu Thr Ile Lys Gly
1505 1510 1515 1520

ggg caa aca gac acc aat aag cta acc gat aat aac atc ggt gtg gta 5149
Gly Gln Thr Asp Thr Asn Lys Leu Thr Asp Asn Asn Ile Gly Val Val
1525 1530 1535

gca ggt act gat ggc ttc act gtc aaa ctt gcc aaa gac cta acc aat 5197
Ala Gly Thr Asp Gly Phe Thr Val Lys Leu Ala Lys Asp Leu Thr Asn
1540 1545 1550

ctt aac agc gtt aat gca ggt ggc acc aaa att gat gac aaa ggc gtg 5245
Leu Asn Ser Val Asn Ala Gly Gly Thr Lys Ile Asp Asp Lys Gly Val
1555 1560 1565

tct ttt gta gac tca agc ggt caa gcc aaa gca aac acc cct gtg cta 5293
Ser Phe Val Asp Ser Ser Gly Gln Ala Lys Ala Asn Thr Pro Val Leu
1570 1575 1580

agt gcc aat ggg ctg gac ctg ggt ggc aag gtc atc agt aat gtg ggc 5341
Ser Ala Asn Gly Leu Asp Leu Gly Gly Lys Val Ile Ser Asn Val Gly
1585 1590 1595 1600

aaa ggc aca aaa gat acc gac gct gcc aat gta caa cag tta aac gaa 5389
Lys Gly Thr Lys Asp Thr Asp Ala Ala Asn Val Gln Gln Leu Asn Glu
1605 1610 1615

gta cgc aac ttg ttg ggt ctt ggt aat gct ggt aat gat aac gct gac 5437
Val Arg Asn Leu Leu Gly Leu Gly Asn Ala Gly Asn Asp Asn Ala Asp
1620 1625 1630

ggc aat cag gta aac att gcc gac atc aaa aaa gac cca aat tca ggt 5485
Gly Asn Gln Val Asn Ile Ala Asp Ile Lys Lys Asp Pro Asn Ser Gly
1635 1640 1645

tca tca tct aac cgc act gtc atc aaa gca ggc acg gta ctt ggc ggt 5533
Ser Ser Ser Asn Arg Thr Val Ile Lys Ala Gly Thr Val Leu Gly Gly
1650 1655 1660

aaa ggt aat aac gat acc gaa aaa ctt gcc act ggt ggt ata caa gtg 5581
Lys Gly Asn Asn Asp Thr Glu Lys Leu Ala Thr Gly Gly Ile Gln Val
1665 1670 1675 1680

ggc gtg gat aaa gac ggc aac gct aac ggc gat tta agc aat gtt tgg 5629
Gly Val Asp Lys Asp Gly Asn Ala Asn Gly Asp Leu Ser Asn Val Trp
1685 1690 1695

gtc aaa acc caa aaa gat ggc agc aaa aaa gcc ctg ctc gcc act tat 5677
Val Lys Thr Gln Lys Asp Gly Ser Lys Lys Ala Leu Leu Ala Thr Tyr
1700 1705 1710

aac gcc gca ggt cag acc aac tat ttg acc aac aac ccc gca gaa gcc 5725
Asn Ala Ala Gly Gln Thr Asn Tyr Leu Thr Asn Asn Pro Ala Glu Ala
1715 1720 1725

att gac aga ata aat gaa caa ggt atc cgc ttc ttc cat gtc aac gat 5773
Ile Asp Arg Ile Asn Glu Gln Gly Ile Arg Phe Phe His Val Asn Asp
1730 1735 1740

ggc aat caa gag cct gtg gta caa ggg cgt aac ggc att gac tca agt 5821
Gly Asn Gln Glu Pro Val Val Gln Gly Arg Asn Gly Ile Asp Ser Ser
1745 1750 1755 1760

gcc tca ggc aag cac tca gtg gcg ata ggt ttc cag gcc aag gca gat 5869
Ala Ser Gly Lys His Ser Val Ala Ile Gly Phe Gln Ala Lys Ala Asp
1765 1770 1775

ggt gaa gcc gcc gtt gcc ata ggc aga caa acc caa gca ggc aac caa 5917
Gly Glu Ala Ala Val Ala Ile Gly Arg Gln Thr Gln Ala Gly Asn Gln
1780 1785 1790

tcc atc gcc atc ggt gat aac gca caa gcc acg ggc gat caa tcc atc 5965
Ser Ile Ala Ile Gly Asp Asn Ala Gln Ala Thr Gly Asp Gln Ser Ile
1795 1800 1805

gcc atc ggt aca ggc aat gtg gta gca ggt aag cac tct ggt gcc atc 6013
Ala Ile Gly Thr Gly Asn Val Val Ala Gly Lys His Ser Gly Ala Ile
1810 1815 1820

ggc gac cca agc act gtt aag gct gat aac agt tac agt gtg ggt aat 6061
Gly Asp Pro Ser Thr Val Lys Ala Asp Asn Ser Tyr Ser Val Gly Asn
1825 1830 1835 1840

aac aac cag ttt acc gat gcc act caa acc gat gtc ttt ggt gtg ggc 6109
Asn Asn Gln Phe Thr Asp Ala Thr Gln Thr Asp Val Phe Gly Val Gly
1845 1850 1855

aat aac atc acc gtg acc gaa agt aac tcg gtt gcc tta ggt tca aac 6157
Asn Asn Ile Thr Val Thr Glu Ser Asn Ser Val Ala Leu Gly Ser Asn
1860 1865 1870

tct gcc atc agt gca ggc aca cac gca ggc aca caa gcc aaa aaa tct 6205
Ser Ala Ile Ser Ala Gly Thr His Ala Gly Thr Gln Ala Lys Lys Ser
1875 1880 1885

gac ggc aca gca ggt aca acc acc aca gca ggt gca acc ggt acg gtt 6253
Asp Gly Thr Ala Gly Thr Thr Thr Thr Ala Gly Ala Thr Gly Thr Val
1890 1895 1900

aaa ggc ttt gct gga caa acg gcg gtt ggt gcg gtc tcc gtg ggt gcc 6301
Lys Gly Phe Ala Gly Gln Thr Ala Val Gly Ala Val Ser Val Gly Ala
1905 1910 1915 1920

tca ggt gct gaa cgc cgt atc caa aat gtg gca gca ggt gag gtc agt 6349
Ser Gly Ala Glu Arg Arg Ile Gln Asn Val Ala Ala Gly Glu Val Ser
1925 1930 1935

gcc acc agc acc gat gcg gtc aat ggt agc cag ttg tac aaa gcc acc 6397
Ala Thr Ser Thr Asp Ala Val Asn Gly Ser Gln Leu Tyr Lys Ala Thr
1940 1945 1950

caa agc att gcc aac gca acc aat gag ctt gac cat cgt atc cac caa 6445
Gln Ser Ile Ala Asn Ala Thr Asn Glu Leu Asp His Arg Ile His Gln
1955 1960 1965

aac gaa aat aag gcc aat gca ggg att tca tca gcg atg gcg atg gcg 6493
Asn Glu Asn Lys Ala Asn Ala Gly Ile Ser Ser Ala Met Ala Met Ala
1970 1975 1980

tcc atg cca caa gcc tac att cct ggc aga tcc atg gtt acc ggg ggt 6541
Ser Met Pro Gln Ala Tyr Ile Pro Gly Arg Ser Met Val Thr Gly Gly
1985 1990 1995 2000

att gcc acc cac aac ggt caa ggt gcg gtg gca gtg gga ctg tcg aag 6589
Ile Ala Thr His Asn Gly Gln Gly Ala Val Ala Val Gly Leu Ser Lys
2005 2010 2015

ctg tcg gat aat ggt caa tgg gta ttt aaa atc aat ggt tca gcc gat 6637
Leu Ser Asp Asn Gly Gln Trp Val Phe Lys Ile Asn Gly Ser Ala Asp
2020 2025 2030

acc caa ggc cat gta ggg gcg gca gtt ggt gca ggt ttt cac ttt 6682
Thr Gln Gly His Val Gly Ala Ala Val Gly Ala Gly Phe His Phe

2035 2040 2045

taagccataa atcgcaagat tttacttaaa aatcaatctc accatagttg tataaaacag 6742

catcagcatc agtcatatta ctgatgctga tgttttttat cacttaaacc attttaccgc 6802

tcaagtgatt ctctttcacc atgaccaaat cgccattgat cataggtaaa cttattgagt 6862

aaattttatc aatgtagttg ttagatatgg ttaaaattgt gccattgacc aaaaaatgac 6922

cgatttatcc cgaaaatttc tgattatgat ccgttgacct gcaggtcgac 6972

Figure 4. *M. catarrhalis* strain Q8 200kDa gene

```
ATG aat cac atc tat aaa gtc atc ttt aac aaa gcc aca ggc aca ttt    48
Met Asn His Ile Tyr Lys Val Ile Phe Asn Lys Ala Thr Gly Thr Phe
  1               5                  10                  15

atg gcc gtg gcg gaa tat gcc aaa tcc cac agt acg ggg ggg ggt agc    96
Met Ala Val Ala Glu Tyr Ala Lys Ser His Ser Thr Gly Gly Gly Ser
             20                  25                  30

tgt gct aca ggg caa gtt ggc agt gta cgc act cta agc ttt gcc cgt   144
Cys Ala Thr Gly Gln Val Gly Ser Val Arg Thr Leu Ser Phe Ala Arg
         35                  40                  45

att gcc gcg ctc gct gtc ctc gtg atc ggt gcg acg ctc aat ggc agt   192
Ile Ala Ala Leu Ala Val Leu Val Ile Gly Ala Thr Leu Asn Gly Ser
     50                  55                  60

gct tat gct caa caa att act acc aag atc gaa att ggt caa aca aac   240
Ala Tyr Ala Gln Gln Ile Thr Thr Lys Ile Glu Ile Gly Gln Thr Asn
 65                  70                  75                  80

aag ata aac aac acg ctg aaa ggc gat gcc cta gcg aca ggt gaa gca   288
Lys Ile Asn Asn Thr Leu Lys Gly Asp Ala Leu Ala Thr Gly Glu Ala
                 85                  90                  95

tcc att gct ttt ggt agt ctt tct aag gca caa ggc tct caa gct att   336
Ser Ile Ala Phe Gly Ser Leu Ser Lys Ala Gln Gly Ser Gln Ala Ile
            100                 105                 110

gct atc ggt agt gtc aaa cca gat cct aat aat ggt agt aat ggt aat   384
Ala Ile Gly Ser Val Lys Pro Asp Pro Asn Asn Gly Ser Asn Gly Asn
        115                 120                 125

gta ggt tcc cac gcc aaa ggt aac gag tcc atc gcc atc ggt ggt gat   432
Val Gly Ser His Ala Lys Gly Asn Glu Ser Ile Ala Ile Gly Gly Asp
    130                 135                 140

gta ttg gct gag ggt gat gcc tcg att gcc atc ggt agt gat gac tta   480
Val Leu Ala Glu Gly Asp Ala Ser Ile Ala Ile Gly Ser Asp Asp Leu
145                 150                 155                 160

tat ttg cct aag aat ctt gat ctg aag aat gaa ttt cac aaa ctt att   528
Tyr Leu Pro Lys Asn Leu Asp Leu Lys Asn Glu Phe His Lys Leu Ile
                165                 170                 175

cat ggc cat gaa ata tta aaa aaa ata caa acc tca acc gat ggt aaa   576
His Gly His Glu Ile Leu Lys Lys Ile Gln Thr Ser Thr Asp Gly Lys
            180                 185                 190

atc aaa tat cga cgc aca aga gca caa ggg cac gcc agt act gca gtg   624
Ile Lys Tyr Arg Arg Thr Arg Ala Gln Gly His Ala Ser Thr Ala Val
        195                 200                 205

gga gcc atg tca tat gca cag ggt cat ttt tcc aac gcc ttt ggt aca   672
Gly Ala Met Ser Tyr Ala Gln Gly His Phe Ser Asn Ala Phe Gly Thr
    210                 215                 220
```

tac gca aca gct gaa gct gcc tat tcc ttg gca gta ggt ctt gcc gcc 720
Tyr Ala Thr Ala Glu Ala Ala Tyr Ser Leu Ala Val Gly Leu Ala Ala
225 230 235 240

caa gcc aca aaa caa tct tca atc gct gtt ggt tcc aat gca aaa gct 768
Gln Ala Thr Lys Gln Ser Ser Ile Ala Val Gly Ser Asn Ala Lys Ala
245 250 255

aac gcg ttt gca gcg aca gcc att ggt gga aat act gta gtt aat ttg 816
Asn Ala Phe Ala Ala Thr Ala Ile Gly Gly Asn Thr Val Val Asn Leu
260 265 270

ggt cga ggc gtt gcc cta ggt ttt ggt tct cag atc ctt gat agg gat 864
Gly Arg Gly Val Ala Leu Gly Phe Gly Ser Gln Ile Leu Asp Arg Asp
275 280 285

aat aat aca gat gcc agt gcc tat gta cca cta ggt aaa acg tta gca 912
Asn Asn Thr Asp Ala Ser Ala Tyr Val Pro Leu Gly Lys Thr Leu Ala
290 295 300

gac cag tat aaa gcc acc cgc cag ggt gat tct acg gat ata ttt tcc 960
Asp Gln Tyr Lys Ala Thr Arg Gln Gly Asp Ser Thr Asp Ile Phe Ser
305 310 315 320

att ggt aat agt aat aat aat aat agc agt atc agg cgt aaa atc atc 1008
Ile Gly Asn Ser Asn Asn Asn Asn Ser Ser Ile Arg Arg Lys Ile Ile
325 330 335

aat gtc ggt gcg ggt tct cgg gat acc gat gcg gtc aat gtg gca cag 1056
Asn Val Gly Ala Gly Ser Arg Asp Thr Asp Ala Val Asn Val Ala Gln
340 345 350

ctt aaa ttg gtg gag gaa ctg gct aat cgt aaa att act ttt aag ggt 1104
Leu Lys Leu Val Glu Glu Leu Ala Asn Arg Lys Ile Thr Phe Lys Gly
355 360 365

gat ggt gac aat aat agc aat agc gta gaa aga ggt ttg ggc aat act 1152
Asp Gly Asp Asn Asn Ser Asn Ser Val Glu Arg Gly Leu Gly Asn Thr
370 375 380

tta act att aaa ggt gat gca cag acc aac gca tta acc gaa gct aac 1200
Leu Thr Ile Lys Gly Asp Ala Gln Thr Asn Ala Leu Thr Glu Ala Asn
385 390 395 400

atc ggt gtg gta aca gat ggc aat ggt ctg aaa gtt aaa ctt gct aaa 1248
Ile Gly Val Val Thr Asp Gly Asn Gly Leu Lys Val Lys Leu Ala Lys
405 410 415

gag ctg act gga ttg acc agt gtc tcc gct acc aac aaa atc acc gtt 1296
Glu Leu Thr Gly Leu Thr Ser Val Ser Ala Thr Asn Lys Ile Thr Val
420 425 430

agt aat acc aac aac aac aac gcc gag cta caa agc ggt ggt ttg acc 1344
Ser Asn Thr Asn Asn Asn Asn Ala Glu Leu Gln Ser Gly Gly Leu Thr
435 440 445

ttt agc cca ata aca ggt aca aaa aca gat aaa acc gtc tac agc att 1392
Phe Ser Pro Ile Thr Gly Thr Lys Thr Asp Lys Thr Val Tyr Ser Ile
450 455 460

```
gat gga ttg aag ttt act aat gat agt aat agt ata gca act aaa ggt    1440
Asp Gly Leu Lys Phe Thr Asn Asp Ser Asn Ser Ile Ala Thr Lys Gly
465                 470                 475                 480

act act cgt att acc aaa aag aaa att ggt ttt gct ggt act aat gat    1488
Thr Thr Arg Ile Thr Lys Lys Lys Ile Gly Phe Ala Gly Thr Asn Asp
                485                 490                 495

gga gtt gat gaa agc aaa cct tat ctt gac aac gaa aag cta aaa gtt    1536
Gly Val Asp Glu Ser Lys Pro Tyr Leu Asp Asn Glu Lys Leu Lys Val
            500                 505                 510

ggc aac agc acc cta aac agt ggt agc ttg act gtt aat aac acc act    1584
Gly Asn Ser Thr Leu Asn Ser Gly Ser Leu Thr Val Asn Asn Thr Thr
        515                 520                 525

ggt aat aaa caa atc caa gtc ggt gct aat ggc att aaa ttt gcc aca    1632
Gly Asn Lys Gln Ile Gln Val Gly Ala Asn Gly Ile Lys Phe Ala Thr
    530                 535                 540

gtc gct aat aat gtt gca aat acc tca gca aca gtc ggc act gct cgt    1680
Val Ala Asn Asn Val Ala Asn Thr Ser Ala Thr Val Gly Thr Ala Arg
545                 550                 555                 560

att acc gaa gag aaa att ggt ttt gct ggt act aat gat gga gtt gat    1728
Ile Thr Glu Glu Lys Ile Gly Phe Ala Gly Thr Asn Asp Gly Val Asp
                565                 570                 575

gaa caa gca cca tat ttg gat aaa gaa cga ctt aaa gtg ggt cgt gtt    1776
Glu Gln Ala Pro Tyr Leu Asp Lys Glu Arg Leu Lys Val Gly Arg Val
            580                 585                 590

gaa att acc aca gat agt ggt att aat gct ggt aat cac aag att acc    1824
Glu Ile Thr Thr Asp Ser Gly Ile Asn Ala Gly Asn His Lys Ile Thr
        595                 600                 605

gga ctt act aat ggt ata gca aat acc gat gcg gtt acc atc aaa cag    1872
Gly Leu Thr Asn Gly Ile Ala Asn Thr Asp Ala Val Thr Ile Lys Gln
    610                 615                 620

ctc aaa gac gcc aag cct act tta aac gca ggc gat ggc atc agt att    1920
Leu Lys Asp Ala Lys Pro Thr Leu Asn Ala Gly Asp Gly Ile Ser Ile
625                 630                 635                 640

aat agt aat aac ggg gat cta gtt gat agt agt ggc aat att acc acc    1968
Asn Ser Asn Asn Gly Asp Leu Val Asp Ser Ser Gly Asn Ile Thr Thr
                645                 650                 655

cca act tat aac att agc gtg aaa acc act aag ctt aac agt aat ggc    2016
Pro Thr Tyr Asn Ile Ser Val Lys Thr Thr Lys Leu Asn Ser Asn Gly
            660                 665                 670

acc agt ggt aat aat aaa ttt agt gtt agt aat gct cat gat aac aat    2064
Thr Ser Gly Asn Asn Lys Phe Ser Val Ser Asn Ala His Asp Asn Asn
        675                 680                 685

agc tta gtt acc gcc aaa gat ttg gca gac tat cta aat aaa gtc aat    2112
Ser Leu Val Thr Ala Lys Asp Leu Ala Asp Tyr Leu Asn Lys Val Asn
```

```
    690                 695                 700

gaa acg gct gac agt gct cta cca agc ttt aaa gtc caa aac ggt gat      2160
Glu Thr Ala Asp Ser Ala Leu Pro Ser Phe Lys Val Gln Asn Gly Asp
705                 710                 715                 720

aat agc aac aac gcc atc acc gtg ggt aaa gat aca aac ggc aag acc      2208
Asn Ser Asn Asn Ala Ile Thr Val Gly Lys Asp Thr Asn Gly Lys Thr
                725                 730                 735

ttc aac acc tta aaa ctc aaa ggt gaa aac ggt gtt aat att acg acc      2256
Phe Asn Thr Leu Lys Leu Lys Gly Glu Asn Gly Val Asn Ile Thr Thr
            740                 745                 750

aat aga gcc aca ggt aca gtt acc ttt ggc att gac caa agt aat ggt      2304
Asn Arg Ala Thr Gly Thr Val Thr Phe Gly Ile Asp Gln Ser Asn Gly
        755                 760                 765

ctc acc acg cct aag ctg acc gtg ggt agc gat aca aat ggt aat cga      2352
Leu Thr Thr Pro Lys Leu Thr Val Gly Ser Asp Thr Asn Gly Asn Arg
    770                 775                 780

ttg gtt att gag caa gtc cct agc gct gac ggt aac agc acc aaa aac      2400
Leu Val Ile Glu Gln Val Pro Ser Ala Asp Gly Asn Ser Thr Lys Asn
785                 790                 795                 800

atc att aaa gga ttg tcc cca aca ctg cct agc att gcc agt cca agt      2448
Ile Ile Lys Gly Leu Ser Pro Thr Leu Pro Ser Ile Ala Ser Pro Ser
                805                 810                 815

ggc cgc aac ata gca ctg ggc aat aca atc gaa gaa aaa gac aaa tcc      2496
Gly Arg Asn Ile Ala Leu Gly Asn Thr Ile Glu Glu Lys Asp Lys Ser
            820                 825                 830

aac gct gcc agc att gat gat gtg cta aat gca ggc ttt aac cta aaa      2544
Asn Ala Ala Ser Ile Asp Asp Val Leu Asn Ala Gly Phe Asn Leu Lys
        835                 840                 845

aat aat ggc aaa gac aaa gac ttt gtc tcc act tat gac act gtt gac      2592
Asn Asn Gly Lys Asp Lys Asp Phe Val Ser Thr Tyr Asp Thr Val Asp
    850                 855                 860

ttt atc gat ggc aat gcc acc acc gcc aca gta act tat gat gaa gcc      2640
Phe Ile Asp Gly Asn Ala Thr Thr Ala Thr Val Thr Tyr Asp Glu Ala
865                 870                 875                 880

aat caa acc agt aaa gtg gcg tat gat gtg aat gtg gat gag aaa acc      2688
Asn Gln Thr Ser Lys Val Ala Tyr Asp Val Asn Val Asp Glu Lys Thr
                885                 890                 895

att gaa ctg aca ggc gat aat ggc aag aaa caa ctt ggc gtc aaa acc      2736
Ile Glu Leu Thr Gly Asp Asn Gly Lys Lys Gln Leu Gly Val Lys Thr
            900                 905                 910

atc aaa ctg acc gaa aca agt act aat ggt aat gca act aca ttt agt      2784
Ile Lys Leu Thr Glu Thr Ser Thr Asn Gly Asn Ala Thr Thr Phe Ser
        915                 920                 925

acc gac gat gac cat gcc ctt gtt aaa gcc agt gat atc gcc ggc aat      2832
```

Thr Asp Asp Asp His Ala Leu Val Lys Ala Ser Asp Ile Ala Gly Asn
930 935 940

cta aac acc cta gcc gag gaa att cac acc acc aaa ggc aca gca aac 2880
Leu Asn Thr Leu Ala Glu Glu Ile His Thr Thr Lys Gly Thr Ala Asn
945 950 955 960

acc gcc cta caa acc ttt acc gtt aaa aag gta gat gaa aat gat aag 2928
Thr Ala Leu Gln Thr Phe Thr Val Lys Lys Val Asp Glu Asn Asp Lys
965 970 975

gct gat gac acc aac gcc atc acc gtg ggt aaa gat ggc aca agt ggt 2976
Ala Asp Asp Thr Asn Ala Ile Thr Val Gly Lys Asp Gly Thr Ser Gly
980 985 990

aaa gtc aac acc tta aaa ctc aaa ggt aaa aac ggt ctt gat att aaa 3024
Lys Val Asn Thr Leu Lys Leu Lys Gly Lys Asn Gly Leu Asp Ile Lys
995 1000 1005

acc gac aaa gat ggt acg gtt acc ttt ggc att aac acc caa agc ggt 3072
Thr Asp Lys Asp Gly Thr Val Thr Phe Gly Ile Asn Thr Gln Ser Gly
1010 1015 1020

ctt aaa gcc ggc gac agc acc act cta aac aac aat ggc ttg tct att 3120
Leu Lys Ala Gly Asp Ser Thr Thr Leu Asn Asn Asn Gly Leu Ser Ile
1025 1030 1035 1040

aaa aac acc gct agt aac gaa caa atc caa gtc ggt gct gat ggc gtg 3168
Lys Asn Thr Ala Ser Asn Glu Gln Ile Gln Val Gly Ala Asp Gly Val
1045 1050 1055

aag ttt gcc atg gtt aat aat ggt gtt gta ggt gct ggc att gat ggc 3216
Lys Phe Ala Met Val Asn Asn Gly Val Val Gly Ala Gly Ile Asp Gly
1060 1065 1070

aca act cgc att acc aga gat gaa att ggc ttt act ggg act aat ggc 3264
Thr Thr Arg Ile Thr Arg Asp Glu Ile Gly Phe Thr Gly Thr Asn Gly
1075 1080 1085

tca ctt gat aaa agc aaa ccc cac cta agc aaa gac ggc att aac gca 3312
Ser Leu Asp Lys Ser Lys Pro His Leu Ser Lys Asp Gly Ile Asn Ala
1090 1095 1100

ggt ggt aaa aag att acc aac att caa tca ggt gag att gcc aaa aac 3360
Gly Gly Lys Lys Ile Thr Asn Ile Gln Ser Gly Glu Ile Ala Lys Asn
1105 1110 1115 1120

agc cat gat gct gtg aca ggc ggc aag att tat gat tta aaa acc gaa 3408
Ser His Asp Ala Val Thr Gly Gly Lys Ile Tyr Asp Leu Lys Thr Glu
1125 1130 1135

ctt gaa aat aaa atc agc agt act gcc aaa aca gca caa aac tca tta 3456
Leu Glu Asn Lys Ile Ser Ser Thr Ala Lys Thr Ala Gln Asn Ser Leu
1140 1145 1150

cac gaa ttc tca gta gca gat gaa caa ggt aat aac ttt acg gtt agt 3504
His Glu Phe Ser Val Ala Asp Glu Gln Gly Asn Asn Phe Thr Val Ser
1155 1160 1165

```
aac cct tac tcc agt tat gac acc tca aag acc tct gat gtc atc acc    3552
Asn Pro Tyr Ser Ser Tyr Asp Thr Ser Lys Thr Ser Asp Val Ile Thr
   1170                1175                1180

ttt gca ggt gaa aac ggc att acc acc aag gta aat aaa ggt gtg gtg    3600
Phe Ala Gly Glu Asn Gly Ile Thr Thr Lys Val Asn Lys Gly Val Val
1185                1190                1195                1200

cgt gtg ggc att gac caa acc aaa ggc tta acc acg cct aag ctg acc    3648
Arg Val Gly Ile Asp Gln Thr Lys Gly Leu Thr Thr Pro Lys Leu Thr
                1205                1210                1215

gtg ggt aat aat aat ggc aaa ggc att gtc att aac agc caa aat ggt    3696
Val Gly Asn Asn Asn Gly Lys Gly Ile Val Ile Asn Ser Gln Asn Gly
            1220                1225                1230

caa aat acc atc aca gga cta agc aac act cta gct aat gtt acc aat    3744
Gln Asn Thr Ile Thr Gly Leu Ser Asn Thr Leu Ala Asn Val Thr Asn
        1235                1240                1245

gat aaa ggt agc gta cgc acc aca gaa cag ggc aat ata atc aaa gac    3792
Asp Lys Gly Ser Val Arg Thr Thr Glu Gln Gly Asn Ile Ile Lys Asp
    1250                1255                1260

gaa gac aaa acc cgt gcc gcc agc att gtt gat gtg cta agc gca ggc    3840
Glu Asp Lys Thr Arg Ala Ala Ser Ile Val Asp Val Leu Ser Ala Gly
1265                1270                1275                1280

ttt aac ttg caa ggc aat ggt gaa gcg gtt gac ttt gtc tcc act tat    3888
Phe Asn Leu Gln Gly Asn Gly Glu Ala Val Asp Phe Val Ser Thr Tyr
                1285                1290                1295

gac acc gtc aac ttt gcc aat ggc aat acc acc acc gct aag gtg acc    3936
Asp Thr Val Asn Phe Ala Asn Gly Asn Thr Thr Thr Ala Lys Val Thr
            1300                1305                1310

tat gat gac aca agc aaa acc agt aaa gtg gtc tat gat gtc aat gtg    3984
Tyr Asp Asp Thr Ser Lys Thr Ser Lys Val Val Tyr Asp Val Asn Val
        1315                1320                1325

gat gat aca acc att gaa gtt aaa gat aaa aaa ctt ggc gta aaa acc    4032
Asp Asp Thr Thr Ile Glu Val Lys Asp Lys Lys Leu Gly Val Lys Thr
    1330                1335                1340

acc aca ttg acc agt act ggc aca ggt gct aat aaa ttt gcc cta agc    4080
Thr Thr Leu Thr Ser Thr Gly Thr Gly Ala Asn Lys Phe Ala Leu Ser
1345                1350                1355                1360

aat caa gct act ggc gat gcg ctt gtc aag gcc agt gat atc gtt gct    4128
Asn Gln Ala Thr Gly Asp Ala Leu Val Lys Ala Ser Asp Ile Val Ala
                1365                1370                1375

cat cta aac acc tta tct ggc gac atc caa act gcc aaa ggg gca agc    4176
His Leu Asn Thr Leu Ser Gly Asp Ile Gln Thr Ala Lys Gly Ala Ser
            1380                1385                1390

caa gcg aac aac tca gca ggc tat gtg gat gct gat ggc aat aag gtc    4224
Gln Ala Asn Asn Ser Ala Gly Tyr Val Asp Ala Asp Gly Asn Lys Val
        1395                1400                1405
```

```
atc tat gac agt acc gat aac aag tac tat caa gcc aaa aat gat ggc     4272
Ile Tyr Asp Ser Thr Asp Asn Lys Tyr Tyr Gln Ala Lys Asn Asp Gly
    1410                1415                1420

aca gtt gat aaa acc aaa gaa gtt gcc aaa gac aaa ctg gtc gcc caa     4320
Thr Val Asp Lys Thr Lys Glu Val Ala Lys Asp Lys Leu Val Ala Gln
1425                1430                1435                1440

gcc caa acc cca gat ggc aca ttg gct caa atg aat gtc aaa tca gtc     4368
Ala Gln Thr Pro Asp Gly Thr Leu Ala Gln Met Asn Val Lys Ser Val
                1445                1450                1455

att aac aaa gaa caa gta aat gat gcc aat aaa aag caa ggc atc aat     4416
Ile Asn Lys Glu Gln Val Asn Asp Ala Asn Lys Lys Gln Gly Ile Asn
            1460                1465                1470

gaa gac aac gcc ttt gtt aaa gga ctt gaa aaa gcc gct tct gat aac     4464
Glu Asp Asn Ala Phe Val Lys Gly Leu Glu Lys Ala Ala Ser Asp Asn
        1475                1480                1485

aaa acc aaa aac gcc gca gta act gtg ggt gat tta aat gcc gtt gcc     4512
Lys Thr Lys Asn Ala Ala Val Thr Val Gly Asp Leu Asn Ala Val Ala
    1490                1495                1500

caa aca ccg ctg acc ttt gca ggg gat aca ggc aca acg gct aaa aaa     4560
Gln Thr Pro Leu Thr Phe Ala Gly Asp Thr Gly Thr Thr Ala Lys Lys
1505                1510                1515                1520

ctg ggc gag act ttg acc atc aaa ggt ggg caa aca gac acc aat aag     4608
Leu Gly Glu Thr Leu Thr Ile Lys Gly Gly Gln Thr Asp Thr Asn Lys
                1525                1530                1535

cta acc gat aat aac atc ggt gtg gta gca ggt act gat ggc ttc act     4656
Leu Thr Asp Asn Asn Ile Gly Val Val Ala Gly Thr Asp Gly Phe Thr
            1540                1545                1550

gtc aaa ctt gcc aaa gac cta acc aat ctt aac agc gtt aat gca ggt     4704
Val Lys Leu Ala Lys Asp Leu Thr Asn Leu Asn Ser Val Asn Ala Gly
        1555                1560                1565

ggc acc aaa att gat gaa aaa ggc atc tct ttt gta gac gca aac ggt     4752
Gly Thr Lys Ile Asp Glu Lys Gly Ile Ser Phe Val Asp Ala Asn Gly
    1570                1575                1580

caa gcc aaa gca aac acc cct gtg cta agt gcc aat ggg ctg gac ctg     4800
Gln Ala Lys Ala Asn Thr Pro Val Leu Ser Ala Asn Gly Leu Asp Leu
1585                1590                1595                1600

ggt ggc aag gtc atc agt aat gtg ggc aaa ggc aca aaa gat acc gac     4848
Gly Gly Lys Val Ile Ser Asn Val Gly Lys Gly Thr Lys Asp Thr Asp
                1605                1610                1615

gct gcc aat gta caa cag tta aac gaa gta cgc aac ttg ttg ggt ctt     4896
Ala Ala Asn Val Gln Gln Leu Asn Glu Val Arg Asn Leu Leu Gly Leu
            1620                1625                1630

ggt aat gat aac gct gac ggc aat cag gta aac att gcc gac atc aaa     4944
Gly Asn Asp Asn Ala Asp Gly Asn Gln Val Asn Ile Ala Asp Ile Lys
```

1635 1640 1645

aaa gac cca aat tca ggt tca tca tct aac cgc act gtc atc aaa gca 4992
Lys Asp Pro Asn Ser Gly Ser Ser Ser Asn Arg Thr Val Ile Lys Ala
1650 1655 1660

ggc acg gta ctt ggc ggt aaa ggt aat aac gat acc gaa aaa ctt gcc 5040
Gly Thr Val Leu Gly Gly Lys Gly Asn Asn Asp Thr Glu Lys Leu Ala
1665 1670 1675 1680

act ggt ggt gta caa gtg ggc gtg gat aaa gac ggc aac gct aac ggc 5088
Thr Gly Gly Val Gln Val Gly Val Asp Lys Asp Gly Asn Ala Asn Gly
1685 1690 1695

gat tta agc aat gtt tgg gtc aaa acc caa aaa gat ggc agc aaa aaa 5136
Asp Leu Ser Asn Val Trp Val Lys Thr Gln Lys Asp Gly Ser Lys Lys
1700 1705 1710

gcc ctg ctc gcc act tat aac gcc gca ggt cag acc aac tat gtg acc 5184
Ala Leu Leu Ala Thr Tyr Asn Ala Ala Gly Gln Thr Asn Tyr Val Thr
1715 1720 1725

aac aac ccc gca gaa gcc att gac aga ata aat gaa caa ggt atc cgc 5232
Asn Asn Pro Ala Glu Ala Ile Asp Arg Ile Asn Glu Gln Gly Ile Arg
1730 1735 1740

ttc ttc cat gtc aac gat ggc aat caa gag cct gtg gta caa ggg cgt 5280
Phe Phe His Val Asn Asp Gly Asn Gln Glu Pro Val Val Gln Gly Arg
1745 1750 1755 1760

aac ggc att gac tca agt gcc tca ggc aag cac tca gtg gcg ata ggt 5328
Asn Gly Ile Asp Ser Ser Ala Ser Gly Lys His Ser Val Ala Ile Gly
1765 1770 1775

ttc cag gcc aag gca gat ggt gaa gcc gcc gtt gcc ata ggc aga caa 5376
Phe Gln Ala Lys Ala Asp Gly Glu Ala Ala Val Ala Ile Gly Arg Gln
1780 1785 1790

acc caa gca ggc aac caa tcc atc gcc atc ggt gat aac gca caa gcc 5424
Thr Gln Ala Gly Asn Gln Ser Ile Ala Ile Gly Asp Asn Ala Gln Ala
1795 1800 1805

acg ggc gat caa tcc atc gcc atc ggt aca ggc aat gtg gta gca ggt 5472
Thr Gly Asp Gln Ser Ile Ala Ile Gly Thr Gly Asn Val Val Ala Gly
1810 1815 1820

aag cac tct ggt gcc atc ggc gac cca agc act gtt aag gct gat aac 5520
Lys His Ser Gly Ala Ile Gly Asp Pro Ser Thr Val Lys Ala Asp Asn
1825 1830 1835 1840

agt tac agt gtg ggt aat aac aac cag ttt acc gat gcc act caa acc 5568
Ser Tyr Ser Val Gly Asn Asn Asn Gln Phe Thr Asp Ala Thr Gln Thr
1845 1850 1855

gat gtc ttt ggt gtg ggc aat aac atc acc gtg acc gaa agt aac tcg 5616
Asp Val Phe Gly Val Gly Asn Asn Ile Thr Val Thr Glu Ser Asn Ser
1860 1865 1870

gtt gcc tta ggt tca aac tct gcc atc agt gca ggc aca cac gca ggc 5664

```
Val Ala Leu Gly Ser Asn Ser Ala Ile Ser Ala Gly Thr His Ala Gly
        1875                1880                1885

aca caa gcc aaa aaa tct gac ggc aca gca ggt aca acc acc aca gca      5712
Thr Gln Ala Lys Lys Ser Asp Gly Thr Ala Gly Thr Thr Thr Thr Ala
    1890                1895                1900

ggt gcc aca ggt acg gtt aaa ggc ttt gct gga caa acg gcg gtt ggt      5760
Gly Ala Thr Gly Thr Val Lys Gly Phe Ala Gly Gln Thr Ala Val Gly
1905                1910                1915                1920

gcg gtc tcc gtg ggt gcc tca ggt gct gaa cgc cgt atc caa aat gtg      5808
Ala Val Ser Val Gly Ala Ser Gly Ala Glu Arg Arg Ile Gln Asn Val
                1925                1930                1935

gca gca ggt gag gtc agt gcc acc agc acc gat gcg gtc aat ggt agc      5856
Ala Ala Gly Glu Val Ser Ala Thr Ser Thr Asp Ala Val Asn Gly Ser
            1940                1945                1950

cag ttg tac aaa gcc acc caa agc att gcc aac gca acc aat gag ctt      5904
Gln Leu Tyr Lys Ala Thr Gln Ser Ile Ala Asn Ala Thr Asn Glu Leu
        1955                1960                1965

gac cat cgt atc cac caa aac gaa aat aaa gcc aat gca ggg att tca      5952
Asp His Arg Ile His Gln Asn Glu Asn Lys Ala Asn Ala Gly Ile Ser
    1970                1975                1980

tca gcg atg gcg atg gcg tcc atg cca caa gcc tac att cct ggc aga      6000
Ser Ala Met Ala Met Ala Ser Met Pro Gln Ala Tyr Ile Pro Gly Arg
1985                1990                1995                2000

tcc atg gtt acc ggg ggt att gcc acc cac aac ggt caa ggt gcg gtg      6048
Ser Met Val Thr Gly Gly Ile Ala Thr His Asn Gly Gln Gly Ala Val
                2005                2010                2015

gca gtg gga ctg tcg aag ctg tcg gat aat ggt caa tgg gta ttt aaa      6096
Ala Val Gly Leu Ser Lys Leu Ser Asp Asn Gly Gln Trp Val Phe Lys
            2020                2025                2030

atc aat ggt tca gcc gat acc caa ggc cat gta ggg gcg gca gtt ggt      6144
Ile Asn Gly Ser Ala Asp Thr Gln Gly His Val Gly Ala Ala Val Gly
        2035                2040                2045

gca ggt ttt cac ttt                                                  6159
Ala Gly Phe His Phe
    2050
```

Figure 5. *Moraxella catarrhalis* les1 200kDa

```
ATG aat cac atc tat aaa gtc atc ttt aac aaa gcc aca ggc aca ttt    48
Met Asn His Ile Tyr Lys Val Ile Phe Asn Lys Ala Thr Gly Thr Phe
  1               5                  10                  15

atg gcc gtg gca gag tgc gcc aaa tcc cac agc gga ggg agt agc agt    96
Met Ala Val Ala Glu Cys Ala Lys Ser His Ser Gly Gly Ser Ser Ser
                 20                  25                  30

agt acc gca gga cag gtg ggc agc tct cct gtc atc cgc ctg act cgt   144
Ser Thr Ala Gly Gln Val Gly Ser Ser Pro Val Ile Arg Leu Thr Arg
             35                  40                  45

gtt gcc acg ctc gct atc cta gtg atc ggt gcg acg ctc aat ggc agt   192
Val Ala Thr Leu Ala Ile Leu Val Ile Gly Ala Thr Leu Asn Gly Ser
     50                  55                  60

gct tat gct caa aat aat agc aag atc gca ttt ggt acc aca ggc aac   240
Ala Tyr Ala Gln Asn Asn Ser Lys Ile Ala Phe Gly Thr Thr Gly Asn
 65                  70                  75                  80

aat gac aat gcc tcg gct agc aat gaa gca tcc att gct att ggt agt   288
Asn Asp Asn Ala Ser Ala Ser Asn Glu Ala Ser Ile Ala Ile Gly Ser
                     85                  90                  95

ctt gct aag gca cat gcc aat caa gct att gct atc ggt ggt agc aaa   336
Leu Ala Lys Ala His Ala Asn Gln Ala Ile Ala Ile Gly Gly Ser Lys
                100                 105                 110

cca gat cct cgt aat caa gcg gct aat cag aag gca ggt tcc cac gcc   384
Pro Asp Pro Arg Asn Gln Ala Ala Asn Gln Lys Ala Gly Ser His Ala
            115                 120                 125

aaa ggt aaa gag tcc atc gcc atc ggt ggt gat gta ctg gct gag ggt   432
Lys Gly Lys Glu Ser Ile Ala Ile Gly Gly Asp Val Leu Ala Glu Gly
    130                 135                 140

gat gcc tcg att gcc att ggt agt gat gac tta tat ttg gat agg aat   480
Asp Ala Ser Ile Ala Ile Gly Ser Asp Asp Leu Tyr Leu Asp Arg Asn
145                 150                 155                 160

agc act aac tct aaa tat cca aat ggt ctt ctt agc act ctt att caa   528
Ser Thr Asn Ser Lys Tyr Pro Asn Gly Leu Leu Ser Thr Leu Ile Gln
                    165                 170                 175

aac cat aca gta tta cgc caa ata cga gac tca aat ggt tct cag aaa   576
Asn His Thr Val Leu Arg Gln Ile Arg Asp Ser Asn Gly Ser Gln Lys
                180                 185                 190

tat aga cgc aca gca gca gaa gga cac gcc agt act gca gtg gga gcc   624
Tyr Arg Arg Thr Ala Ala Glu Gly His Ala Ser Thr Ala Val Gly Ala
        195                 200                 205

atg gca tat gca aag ggt cat ttt gcc aac gcc ttt ggt aca cgg tca   672
Met Ala Tyr Ala Lys Gly His Phe Ala Asn Ala Phe Gly Thr Arg Ser
```

210 215 220

aca gct gaa ggc aac tat tcc ttg gca gta ggt ctt acc gcc aaa gcc 720
Thr Ala Glu Gly Asn Tyr Ser Leu Ala Val Gly Leu Thr Ala Lys Ala
225 230 235 240

gaa aaa gga tat aca atc gct att ggt tct aat gca caa gct atc aat 768
Glu Lys Gly Tyr Thr Ile Ala Ile Gly Ser Asn Ala Gln Ala Ile Asn
245 250 255

tat gga gca cta gcc ctt ggt gca gat act cga gtt gat ttg gat tac 816
Tyr Gly Ala Leu Ala Leu Gly Ala Asp Thr Arg Val Asp Leu Asp Tyr
260 265 270

ggt att gcc cta ggt tat ggt tct cag atc ctt aat aat aat aat aat 864
Gly Ile Ala Leu Gly Tyr Gly Ser Gln Ile Leu Asn Asn Asn Asn Asn
275 280 285

aat aat aat aaa gcc tat gta cca gaa ggt aat ggg tca aac ata aaa 912
Asn Asn Asn Lys Ala Tyr Val Pro Glu Gly Asn Gly Ser Asn Ile Lys
290 295 300

tcg tct aaa gcc acc ggc aat ggt tta ttt tcc att ggt agt agc act 960
Ser Ser Lys Ala Thr Gly Asn Gly Leu Phe Ser Ile Gly Ser Ser Thr
305 310 315 320

atc aag cgt aaa atc atc aat gtc ggt gca ggt tat gag gat acc gat 1008
Ile Lys Arg Lys Ile Ile Asn Val Gly Ala Gly Tyr Glu Asp Thr Asp
325 330 335

gcg gtc aat gtg gca cag cta aaa gcg gtg gag aat ctg gct aag cgt 1056
Ala Val Asn Val Ala Gln Leu Lys Ala Val Glu Asn Leu Ala Lys Arg
340 345 350

caa att act ttt aag ggt gat gat aac ggt act ggc gtt aag aaa aaa 1104
Gln Ile Thr Phe Lys Gly Asp Asp Asn Gly Thr Gly Val Lys Lys Lys
355 360 365

ctg ggc gag act tta acc att aaa ggt ggt gag acc caa gcg gac aag 1152
Leu Gly Glu Thr Leu Thr Ile Lys Gly Gly Glu Thr Gln Ala Asp Lys
370 375 380

cta acc gat aat aat aac att ggt gtg gta aca gat aat aat act ggt 1200
Leu Thr Asp Asn Asn Asn Ile Gly Val Val Thr Asp Asn Asn Thr Gly
385 390 395 400

ctg aaa gtt aaa ctt gct aaa aac cta agc ggt ctt gaa aca gtt agc 1248
Leu Lys Val Lys Leu Ala Lys Asn Leu Ser Gly Leu Glu Thr Val Ser
405 410 415

acc aaa aac cta acc gcc agc gag aaa gtt acg gta ggt agt ggt aat 1296
Thr Lys Asn Leu Thr Ala Ser Glu Lys Val Thr Val Gly Ser Gly Asn
420 425 430

aac acc gct gag cta caa agc ggt ggt tta acc ttt acc cca aca aca 1344
Asn Thr Ala Glu Leu Gln Ser Gly Gly Leu Thr Phe Thr Pro Thr Thr
435 440 445

```
aat gca agc aca gac aaa acc gtc tat ggc act gat ggg ctt aag ttt     1392
Asn Ala Ser Thr Asp Lys Thr Val Tyr Gly Thr Asp Gly Leu Lys Phe
    450                 455                 460

act gat aat tct aat acg gca ctt gaa gat act act cgt atc acc aaa     1440
Thr Asp Asn Ser Asn Thr Ala Leu Glu Asp Thr Thr Arg Ile Thr Lys
465                 470                 475                 480

gat aaa att ggt ttt agc aat aaa gct ggt aca gtt gat gaa aac aaa     1488
Asp Lys Ile Gly Phe Ser Asn Lys Ala Gly Thr Val Asp Glu Asn Lys
                485                 490                 495

cct tat ctt gat aaa gac aag cta aaa gtt ggc aac agc acc cta aac     1536
Pro Tyr Leu Asp Lys Asp Lys Leu Lys Val Gly Asn Ser Thr Leu Asn
            500                 505                 510

aac ggt ggc ttg act gtt aat aac acc att ggt ggt agc aat aaa caa     1584
Asn Gly Gly Leu Thr Val Asn Asn Thr Ile Gly Gly Ser Asn Lys Gln
        515                 520                 525

atc caa gtc ggt gct gat ggc att aaa ttt gcc gat gtg aat gtt aat     1632
Ile Gln Val Gly Ala Asp Gly Ile Lys Phe Ala Asp Val Asn Val Asn
    530                 535                 540

gta tca aat gcc gca aaa ttc ggc act act cgt att acc gaa gag gaa     1680
Val Ser Asn Ala Ala Lys Phe Gly Thr Thr Arg Ile Thr Glu Glu Glu
545                 550                 555                 560

att ggc ttt gct gat gct gat ggt aaa gtt gat aaa aag tca cca tat     1728
Ile Gly Phe Ala Asp Ala Asp Gly Lys Val Asp Lys Lys Ser Pro Tyr
                565                 570                 575

ttg gat aaa aaa caa ctt caa gtg ggt ggt gtt aaa att acc aaa gac     1776
Leu Asp Lys Lys Gln Leu Gln Val Gly Gly Val Lys Ile Thr Lys Asp
            580                 585                 590

agt ggc att aat gca ggt gat caa aag atc agt aat gtt aaa gat gca     1824
Ser Gly Ile Asn Ala Gly Asp Gln Lys Ile Ser Asn Val Lys Asp Ala
        595                 600                 605

acg gac gat acc gat gca gtc act tat aaa cag ctt aaa caa gtc caa     1872
Thr Asp Asp Thr Asp Ala Val Thr Tyr Lys Gln Leu Lys Gln Val Gln
    610                 615                 620

caa gac gcc gac ggt gcc cta caa agc ttc tct att cgt gat gaa aaa     1920
Gln Asp Ala Asp Gly Ala Leu Gln Ser Phe Ser Ile Arg Asp Glu Lys
625                 630                 635                 640

ggt cag gaa ttt acg att agt aac ttg tat tct aat ggt aat acc cca     1968
Gly Gln Glu Phe Thr Ile Ser Asn Leu Tyr Ser Asn Gly Asn Thr Pro
                645                 650                 655

aat acc ttt gag acc atc acc ttt gca ggt gaa aac ggc atc agt atc     2016
Asn Thr Phe Glu Thr Ile Thr Phe Ala Gly Glu Asn Gly Ile Ser Ile
            660                 665                 670
```

agc aat gac ata gcc aaa ggt aaa gtc aaa gtt ggt att gac cca atc 2064
Ser Asn Asp Ile Ala Lys Gly Lys Val Lys Val Gly Ile Asp Pro Ile
675 680 685

aat ggt ctc acc acg cct aag ctg acc gtg ggt agc gat aaa gat ggt 2112
Asn Gly Leu Thr Thr Pro Lys Leu Thr Val Gly Ser Asp Lys Asp Gly
690 695 700

aaa act caa ttg gtt att gag caa gtg gct agc ggt aac gac acc aaa 2160
Lys Thr Gln Leu Val Ile Glu Gln Val Ala Ser Gly Asn Asp Thr Lys
705 710 715 720

aac atc att aga gga ttg tcc cca aca ctg cct agc att acc aat gca 2208
Asn Ile Ile Arg Gly Leu Ser Pro Thr Leu Pro Ser Ile Thr Asn Ala
725 730 735

ggt ggc gta cgc acc aca gaa cag ggc aat aca atc acc agc gac gaa 2256
Gly Gly Val Arg Thr Thr Glu Gln Gly Asn Thr Ile Thr Ser Asp Glu
740 745 750

gac aaa tcc aaa gcc gcc agt atc ggt gat ata tta aat aca ggc ttt 2304
Asp Lys Ser Lys Ala Ala Ser Ile Gly Asp Ile Leu Asn Thr Gly Phe
755 760 765

aac cta aaa aat aat agc aac tcc gtt ggc ttt gtc tcc act tat aac 2352
Asn Leu Lys Asn Asn Ser Asn Ser Val Gly Phe Val Ser Thr Tyr Asn
770 775 780

act gtt gac ttt atc gat ggc aat gcc acc acc gct aag gta act tac 2400
Thr Val Asp Phe Ile Asp Gly Asn Ala Thr Thr Ala Lys Val Thr Tyr
785 790 795 800

gat gaa acc aat caa acc agt aaa gta act tat gat gtc aat gtg gat 2448
Asp Glu Thr Asn Gln Thr Ser Lys Val Thr Tyr Asp Val Asn Val Asp
805 810 815

gag aaa acc att gaa ctc aca ggc gat aat ggc aag aca aac aaa att 2496
Glu Lys Thr Ile Glu Leu Thr Gly Asp Asn Gly Lys Thr Asn Lys Ile
820 825 830

ggc gtc aaa acc acc aca ctg acc aca aca aat gct aat ggt aaa gca 2544
Gly Val Lys Thr Thr Thr Leu Thr Thr Thr Asn Ala Asn Gly Lys Ala
835 840 845

acc aac ttt agt acc acc gat aac gat gcc ctt gtt aac gcc aaa gac 2592
Thr Asn Phe Ser Thr Thr Asp Asn Asp Ala Leu Val Asn Ala Lys Asp
850 855 860

atc gcc gaa aat cta aac acc cta gcc aag gaa att cac acc acc aaa 2640
Ile Ala Glu Asn Leu Asn Thr Leu Ala Lys Glu Ile His Thr Thr Lys
865 870 875 880

ggc aca gca gac acc gcc cta caa acc ttt aaa gtc aaa aaa gac ggt 2688
Gly Thr Ala Asp Thr Ala Leu Gln Thr Phe Lys Val Lys Lys Asp Gly
885 890 895

gca act gat gac gaa acc atc acc gtg ggt aaa gat ggt aca caa aac 2736

Ala Thr Asp Asp Glu Thr Ile Thr Val Gly Lys Asp Gly Thr Gln Asn
900 905 910

ggc aag acc gtc aac act cta aaa ctc aaa ggt gaa aac ggt cta acg 2784
Gly Lys Thr Val Asn Thr Leu Lys Leu Lys Gly Glu Asn Gly Leu Thr
915 920 925

gtt gct acc aat aaa gat ggt acg gtt acc ttt ggc att aac acc caa 2832
Val Ala Thr Asn Lys Asp Gly Thr Val Thr Phe Gly Ile Asn Thr Gln
930 935 940

agc ggt ctt aaa gcc ggc gac agc acc act cta aac aaa gat ggc ttg 2880
Ser Gly Leu Lys Ala Gly Asp Ser Thr Thr Leu Asn Lys Asp Gly Leu
945 950 955 960

tct att aaa aac ccc gct agt aac gaa caa atc caa gtc ggt gct gat 2928
Ser Ile Lys Asn Pro Ala Ser Asn Glu Gln Ile Gln Val Gly Ala Asp
965 970 975

ggc gtg aag ttt gcc aag gtt gat aag ggt aat tca agc act ggc att 2976
Gly Val Lys Phe Ala Lys Val Asp Lys Gly Asn Ser Ser Thr Gly Ile
980 985 990

gat ggc aca agc cgt atc acc aaa gat caa att ggc ttt act ggg gct 3024
Asp Gly Thr Ser Arg Ile Thr Lys Asp Gln Ile Gly Phe Thr Gly Ala
995 1000 1005

aat ggc tca ctt gat acc acc aaa ccc cac cta acc aaa gac aag ctt 3072
Asn Gly Ser Leu Asp Thr Thr Lys Pro His Leu Thr Lys Asp Lys Leu
1010 1015 1020

aaa gtg ggt gaa gtt gaa att acc aac act ggc att aac gca ggt ggt 3120
Lys Val Gly Glu Val Glu Ile Thr Asn Thr Gly Ile Asn Ala Gly Gly
1025 1030 1035 1040

aaa aag att acc aac att caa tca ggt gat att acc caa aac agc aat 3168
Lys Lys Ile Thr Asn Ile Gln Ser Gly Asp Ile Thr Gln Asn Ser Asn
1045 1050 1055

gat gct gtg aca ggc ggt cgg gtt tat gat tta aaa acc gaa ctt gaa 3216
Asp Ala Val Thr Gly Gly Arg Val Tyr Asp Leu Lys Thr Glu Leu Glu
1060 1065 1070

agc aaa atc aac agt gct gct aaa aca gca caa aac tca tta cac gaa 3264
Ser Lys Ile Asn Ser Ala Ala Lys Thr Ala Gln Asn Ser Leu His Glu
1075 1080 1085

ttc tca gta gca gat gaa caa ggt aat cac ttt acg gtt agt aac cct 3312
Phe Ser Val Ala Asp Glu Gln Gly Asn His Phe Thr Val Ser Asn Pro
1090 1095 1100

tac tcc agt tat gac acc tca aag acc tct gat gtc atc acc ttt gca 3360
Tyr Ser Ser Tyr Asp Thr Ser Lys Thr Ser Asp Val Ile Thr Phe Ala
1105 1110 1115 1120

ggt gaa aac ggc att acc acc aag gta aat aaa ggt gtg gtg cgt gtg 3408
Gly Glu Asn Gly Ile Thr Thr Lys Val Asn Lys Gly Val Val Arg Val

1125 1130 1135

ggc att gac caa acc aaa ggc tta acc acg cct aag ctg acc gtg ggt 3456
Gly Ile Asp Gln Thr Lys Gly Leu Thr Thr Pro Lys Leu Thr Val Gly
1140 1145 1150

aat aat aat ggc aaa ggc att gtc att gac agt aaa gat ggt caa aat 3504
Asn Asn Asn Gly Lys Gly Ile Val Ile Asp Ser Lys Asp Gly Gln Asn
1155 1160 1165

acc atc aca gga cta agc aac act cta gct aat gtt acc aat gat ggt 3552
Thr Ile Thr Gly Leu Ser Asn Thr Leu Ala Asn Val Thr Asn Asp Gly
1170 1175 1180

gca gga cac gca cta agc caa ggg ctt gcc aat gac acc gac aaa acc 3600
Ala Gly His Ala Leu Ser Gln Gly Leu Ala Asn Asp Thr Asp Lys Thr
1185 1190 1195 1200

cgt gcc gcc agc att ggt gat gtg cta aac gca ggc ttt aac ttg caa 3648
Arg Ala Ala Ser Ile Gly Asp Val Leu Asn Ala Gly Phe Asn Leu Gln
1205 1210 1215

ggc aat ggt gaa gcg gtt gac ttt gtc tcc act tat gac act gtt gac 3696
Gly Asn Gly Glu Ala Val Asp Phe Val Ser Thr Tyr Asp Thr Val Asp
1220 1225 1230

ttt atc gat ggc aat gcc acc acc gct aag gtg acc tat gat gac aca 3744
Phe Ile Asp Gly Asn Ala Thr Thr Ala Lys Val Thr Tyr Asp Asp Thr
1235 1240 1245

agc aaa acc agt aaa gtg gtc tat gat gtc aat gtg gat aat aaa acc 3792
Ser Lys Thr Ser Lys Val Val Tyr Asp Val Asn Val Asp Asn Lys Thr
1250 1255 1260

att gaa gtg aca agt gat aaa aaa ctt ggc gtc aaa acc acc aca ctg 3840
Ile Glu Val Thr Ser Asp Lys Lys Leu Gly Val Lys Thr Thr Thr Leu
1265 1270 1275 1280

acc aaa aca agt gct aat ggt aat gca acc aaa ttt agt gcc gcc gat 3888
Thr Lys Thr Ser Ala Asn Gly Asn Ala Thr Lys Phe Ser Ala Ala Asp
1285 1290 1295

ggc gat gcc ctt gtt aaa gcc agt gat atc gcc acc cat cta aat acc 3936
Gly Asp Ala Leu Val Lys Ala Ser Asp Ile Ala Thr His Leu Asn Thr
1300 1305 1310

ttg gct ggc gac atc caa acc gcc aaa ggg gca agc caa gca agc agc 3984
Leu Ala Gly Asp Ile Gln Thr Ala Lys Gly Ala Ser Gln Ala Ser Ser
1315 1320 1325

tca gca agc tat gtg gat gct gat ggc aac aag gtc atc tat gac agt 4032
Ser Ala Ser Tyr Val Asp Ala Asp Gly Asn Lys Val Ile Tyr Asp Ser
1330 1335 1340

acc gat aag aag tac tat caa gtc aat gac aag ggt caa gtg gac aaa 4080
Thr Asp Lys Lys Tyr Tyr Gln Val Asn Asp Lys Gly Gln Val Asp Lys
1345 1350 1355 1360

aac aaa gaa gtt gcc aaa gac aaa ctg gtc gcc caa gcc caa acc cca 4128
Asn Lys Glu Val Ala Lys Asp Lys Leu Val Ala Gln Ala Gln Thr Pro
1365 1370 1375

gat ggc aca ttg gct caa atg aat gtc aaa tca gtc att aac aaa gag 4176
Asp Gly Thr Leu Ala Gln Met Asn Val Lys Ser Val Ile Asn Lys Glu
1380 1385 1390

caa gta aat gat gcc aat aaa aag caa ggc atc aat gaa gac aac gcc 4224
Gln Val Asn Asp Ala Asn Lys Lys Gln Gly Ile Asn Glu Asp Asn Ala
1395 1400 1405

ttt atc aaa ggg ctt gaa aac gcc gcc aaa gac acc aaa acc aaa aac 4272
Phe Ile Lys Gly Leu Glu Asn Ala Ala Lys Asp Thr Lys Thr Lys Asn
1410 1415 1420

gcc gca gta act gtg ggt gat tta aat gcc gtt gcc caa aca ccg ctg 4320
Ala Ala Val Thr Val Gly Asp Leu Asn Ala Val Ala Gln Thr Pro Leu
1425 1430 1435 1440

acc ttt gca ggg gat aca ggc aca acg gct aaa aaa ctg ggc gag act 4368
Thr Phe Ala Gly Asp Thr Gly Thr Thr Ala Lys Lys Leu Gly Glu Thr
1445 1450 1455

ttg acc atc aaa ggt ggg caa aca gac acc aat aag cta acc gat aat 4416
Leu Thr Ile Lys Gly Gly Gln Thr Asp Thr Asn Lys Leu Thr Asp Asn
1460 1465 1470

aac atc ggt gtg gta gca ggt act gat ggc ttc act gtc aaa ctt gcc 4464
Asn Ile Gly Val Val Ala Gly Thr Asp Gly Phe Thr Val Lys Leu Ala
1475 1480 1485

aaa gac cta acc aat ctt aac agc gtt aat gca ggt ggc acc aga att 4512
Lys Asp Leu Thr Asn Leu Asn Ser Val Asn Ala Gly Gly Thr Arg Ile
1490 1495 1500

gat gaa aaa ggc atc tct ttt gta gac gca aac ggt caa gcc aaa gca 4560
Asp Glu Lys Gly Ile Ser Phe Val Asp Ala Asn Gly Gln Ala Lys Ala
1505 1510 1515 1520

aac acc cct gtg cta agt gcc aat ggg ctg gac ctg ggt ggc aaa cgc 4608
Asn Thr Pro Val Leu Ser Ala Asn Gly Leu Asp Leu Gly Gly Lys Arg
1525 1530 1535

atc agt aac atc ggt gca gct gtt gat gat aac gat gcg gtg aac ttt 4656
Ile Ser Asn Ile Gly Ala Ala Val Asp Asp Asn Asp Ala Val Asn Phe
1540 1545 1550

aag cag ttt aat gaa gtt gcc aaa acg gtc aac aac cta aac aac caa 4704
Lys Gln Phe Asn Glu Val Ala Lys Thr Val Asn Asn Leu Asn Asn Gln
1555 1560 1565

agt aac tca ggt gcg tca tta ccc ttt gtg gta acc gat gcc aat ggc 4752
Ser Asn Ser Gly Ala Ser Leu Pro Phe Val Val Thr Asp Ala Asn Gly
1570 1575 1580

```
aag ccc atc aat ggc acc gat ggc aag ccc caa aaa gcc atc aag ggc    4800
Lys Pro Ile Asn Gly Thr Asp Gly Lys Pro Gln Lys Ala Ile Lys Gly
1585                1590                1595                1600

gcc gat ggt aaa tac tat cac gcc aac gcc aac ggc gta cct gtg gac    4848
Ala Asp Gly Lys Tyr Tyr His Ala Asn Ala Asn Gly Val Pro Val Asp
                1605                1610                1615

aaa gat ggc aag ccc atc acc gat gcg gac aaa ctt gcc aat ctg gca    4896
Lys Asp Gly Lys Pro Ile Thr Asp Ala Asp Lys Leu Ala Asn Leu Ala
            1620                1625                1630

gct cat ggc aaa ccc ctt gat gca ggt cat caa gtg gtg gca agc cta    4944
Ala His Gly Lys Pro Leu Asp Ala Gly His Gln Val Val Ala Ser Leu
        1635                1640                1645

ggc ggc aac tca gat gcc atc acc cta acc aac atc aag tcc act ttg    4992
Gly Gly Asn Ser Asp Ala Ile Thr Leu Thr Asn Ile Lys Ser Thr Leu
    1650                1655                1660

cca caa att gac aca cca aac aca ggt aat gcc aat gca ggg caa gcc    5040
Pro Gln Ile Asp Thr Pro Asn Thr Gly Asn Ala Asn Ala Gly Gln Ala
1665                1670                1675                1680

caa agt ctg ccc agc cta tca gca gca cag caa agt aat gct gcc agt    5088
Gln Ser Leu Pro Ser Leu Ser Ala Ala Gln Gln Ser Asn Ala Ala Ser
                1685                1690                1695

gtc aaa gat gtg cta aat gta ggc ttt aac ttg cag acc aat cac aat    5136
Val Lys Asp Val Leu Asn Val Gly Phe Asn Leu Gln Thr Asn His Asn
            1700                1705                1710

caa gtg gac ttt gtc aaa gcc tat gat acc gtc aac ttt gtc aat ggt    5184
Gln Val Asp Phe Val Lys Ala Tyr Asp Thr Val Asn Phe Val Asn Gly
        1715                1720                1725

aca ggt gcc gac atc aca agc gtg cgt agt gct gat ggc acg atg agt    5232
Thr Gly Ala Asp Ile Thr Ser Val Arg Ser Ala Asp Gly Thr Met Ser
    1730                1735                1740

aac atc acc gtc aac acc gcc tta gca gcg acc gat gat gat ggc aat    5280
Asn Ile Thr Val Asn Thr Ala Leu Ala Ala Thr Asp Asp Asp Gly Asn
1745                1750                1755                1760

gtg ctt atc aaa gcc aaa gat ggt aag ttc tac aaa gca gac gac ctc    5328
Val Leu Ile Lys Ala Lys Asp Gly Lys Phe Tyr Lys Ala Asp Asp Leu
                1765                1770                1775

atg cca aac ggc tca cta aaa gca ggc aaa tca gcc agt gat gcc aaa    5376
Met Pro Asn Gly Ser Leu Lys Ala Gly Lys Ser Ala Ser Asp Ala Lys
            1780                1785                1790

act cca act ggt cta agc ctt gtt aac ccc aat gct ggt aaa ggc agt    5424
Thr Pro Thr Gly Leu Ser Leu Val Asn Pro Asn Ala Gly Lys Gly Ser
        1795                1800                1805

aca ggc gat gca gtg gct ctt aat aac tta tca aaa gcg gta ttt aaa    5472
```

```
Thr Gly Asp Ala Val Ala Leu Asn Asn Leu Ser Lys Ala Val Phe Lys
   1810                1815                1820

tcc aaa gat ggt aca act act acc aca gta agc tct gat ggc atc agt    5520
Ser Lys Asp Gly Thr Thr Thr Thr Thr Val Ser Ser Asp Gly Ile Ser
1825                1830                1835                1840

atc caa ggc aaa gat aac agc agc atc acc cta agc aaa gat ggg ctg    5568
Ile Gln Gly Lys Asp Asn Ser Ser Ile Thr Leu Ser Lys Asp Gly Leu
                1845                1850                1855

aat gta ggc ggt aag gtc atc agc aat gtg ggt aaa ggc aca aaa gac    5616
Asn Val Gly Gly Lys Val Ile Ser Asn Val Gly Lys Gly Thr Lys Asp
            1860                1865                1870

acc gac gct gcc aat gta caa cag tta aac gaa gta cgc aac ttg ttg    5664
Thr Asp Ala Ala Asn Val Gln Gln Leu Asn Glu Val Arg Asn Leu Leu
        1875                1880                1885

ggt ctt ggt aat gct ggt aat gat aac gct gac ggc aat cag gta aac    5712
Gly Leu Gly Asn Ala Gly Asn Asp Asn Ala Asp Gly Asn Gln Val Asn
   1890                1895                1900

att gcc gac atc aaa aaa gac cca aat tca ggt tca tca tct aac cgc    5760
Ile Ala Asp Ile Lys Lys Asp Pro Asn Ser Gly Ser Ser Ser Asn Arg
1905                1910                1915                1920

act gtc atc aaa gca ggc acg gta ctt ggc ggt aaa ggt aat aac gat    5808
Thr Val Ile Lys Ala Gly Thr Val Leu Gly Gly Lys Gly Asn Asn Asp
                1925                1930                1935

acc gaa aaa ctt gcc act ggt ggt gta caa gtg ggc gtg gat aaa gac    5856
Thr Glu Lys Leu Ala Thr Gly Gly Val Gln Val Gly Val Asp Lys Asp
            1940                1945                1950

ggc aac gct aac ggc gat tta agc aat gtt tgg gtc aaa acc caa aaa    5904
Gly Asn Ala Asn Gly Asp Leu Ser Asn Val Trp Val Lys Thr Gln Lys
        1955                1960                1965

gat ggc agc aaa aaa gcc ctg ctc gcc act tat aac gcc gca ggt cag    5952
Asp Gly Ser Lys Lys Ala Leu Leu Ala Thr Tyr Asn Ala Ala Gly Gln
   1970                1975                1980

acc aac tat ttg acc aac aac ccc gca gaa gcc att gac aga ata aat    6000
Thr Asn Tyr Leu Thr Asn Asn Pro Ala Glu Ala Ile Asp Arg Ile Asn
1985                1990                1995                2000

gaa caa ggt atc cgc ttc ttc cat gtc aac gat ggc aat caa gag cct    6048
Glu Gln Gly Ile Arg Phe Phe His Val Asn Asp Gly Asn Gln Glu Pro
                2005                2010                2015

gtg gta caa ggg cgt aac ggc att gac tca agt gcc tca ggc aag cac    6096
Val Val Gln Gly Arg Asn Gly Ile Asp Ser Ser Ala Ser Gly Lys His
            2020                2025                2030

tca gtg gcg ata ggt ttc cag gcc aag gca gat ggt gaa gcc gcc gtt    6144
Ser Val Ala Ile Gly Phe Gln Ala Lys Ala Asp Gly Glu Ala Ala Val
```

2035 2040 2045

gcc ata ggc aga caa acc caa gca ggc aac caa tcc atc gcc atc ggt 6192
Ala Ile Gly Arg Gln Thr Gln Ala Gly Asn Gln Ser Ile Ala Ile Gly
2050 2055 2060

gat aac gca caa gcc acg ggc gat caa tcc atc gcc atc ggt aca ggc 6240
Asp Asn Ala Gln Ala Thr Gly Asp Gln Ser Ile Ala Ile Gly Thr Gly
2065 2070 2075 2080

aat gtg gta aca ggt aag cac tct ggt gcc atc ggc gac cca agc act 6288
Asn Val Val Thr Gly Lys His Ser Gly Ala Ile Gly Asp Pro Ser Thr
2085 2090 2095

gtt aag gct gat aac agt tac agt gtg ggt aat aac aac cag ttt atc 6336
Val Lys Ala Asp Asn Ser Tyr Ser Val Gly Asn Asn Asn Gln Phe Ile
2100 2105 2110

gat gcc act cag acc gat gtc ttt ggt gtg ggc aat aac atc acc gtg 6384
Asp Ala Thr Gln Thr Asp Val Phe Gly Val Gly Asn Asn Ile Thr Val
2115 2120 2125

acc gaa agt aac tcg gtt gcc tta ggt tca aac tct gcc atc agt gca 6432
Thr Glu Ser Asn Ser Val Ala Leu Gly Ser Asn Ser Ala Ile Ser Ala
2130 2135 2140

ggc aca cac gca ggc aca caa gcc aaa aaa tct gac ggc aca gca ggt 6480
Gly Thr His Ala Gly Thr Gln Ala Lys Lys Ser Asp Gly Thr Ala Gly
2145 2150 2155 2160

aca acc acc aca gca ggt gca aca ggt acg gtt aaa ggc ttt gct gga 6528
Thr Thr Thr Thr Ala Gly Ala Thr Gly Thr Val Lys Gly Phe Ala Gly
2165 2170 2175

caa acg gcg gtt ggt gcg gtc tcc gtg ggt gcc tca ggt gct gaa cgc 6576
Gln Thr Ala Val Gly Ala Val Ser Val Gly Ala Ser Gly Ala Glu Arg
2180 2185 2190

cgt atc caa aat gtg gca gca ggt gag gtc agt gcc acc agc acc gat 6624
Arg Ile Gln Asn Val Ala Ala Gly Glu Val Ser Ala Thr Ser Thr Asp
2195 2200 2205

gcg gtc aat ggt agc cag ttg tac aaa gcc acc caa ggc att gcc aac 6672
Ala Val Asn Gly Ser Gln Leu Tyr Lys Ala Thr Gln Gly Ile Ala Asn
2210 2215 2220

gca acc aat gag ctt gac cat cgt atc cac caa aac gaa aat aaa gcc 6720
Ala Thr Asn Glu Leu Asp His Arg Ile His Gln Asn Glu Asn Lys Ala
2225 2230 2235 2240

aat gca ggg att tca tca gcg atg gcg atg gcg tcc atg cca caa gcc 6768
Asn Ala Gly Ile Ser Ser Ala Met Ala Met Ala Ser Met Pro Gln Ala
2245 2250 2255

tac att cct ggc aga tcc atg gtt acc ggg ggt att gcc acc cac aac 6816
Tyr Ile Pro Gly Arg Ser Met Val Thr Gly Gly Ile Ala Thr His Asn
2260 2265 2270

ggt caa ggt gcg gtg gca gtg gga ctg tcg aag ctg tcg gat aat ggt 6864
Gly Gln Gly Ala Val Ala Val Gly Leu Ser Lys Leu Ser Asp Asn Gly
2275 2280 2285

caa tgg gta ttt aaa atc aat ggt tca gcc gat acc caa ggc cat gta 6912
Gln Trp Val Phe Lys Ile Asn Gly Ser Ala Asp Thr Gln Gly His Val
2290 2295 2300

ggg gcg gca gtt ggt gca ggt ttt cac ttt 6942
Gly Ala Ala Val Gly Ala Gly Phe His Phe
2305 2310

Figure 6. Alignment of amino acid sequence of 200kDa proteins of *M. catarrhalis* strains

```
        10        20        30        40        50        60        70        80         90        100
MNHIYKVIFNKATGTFMAVAEYAKSHSTGGGSCATGQVGSVCTLSFARIAALAVLVIGATLSGSAYAQKKDTKHIAIGEQNQPRRS--GTAKADGDRAIAIG    4223
........................................R..................N......QIT..-.E..QT.KINNTLK.D.L.T.EAS..F.    Q8
....................C.....G.SS.STA.....SPVIRLT.V.T..I.......N......NNSK--..F.TTGNNDNA-----S.SNEAS....    LES-1

       110       120       130       140       150       160       170       180       190       200
ENANAQGGQAIAIGSSNKTVNGSSLD-KIGTDATGQESIAIGGDVKASGDASIAIGSDDLHLLDQHGNPKHPKGTLINDLINGHAVLKEIRSSKDNDVKYR    4223
SLSK...S.......VKPDP.NG.NG-NV.SH.K.N.........L.E............Y.PKNLDL-.NEFHK----..H..EI..K.QT.T.GKI...    Q8
SL.K.HAN......G.KPDPRNQAANQ.A.SH.K.K.........L.E............Y.DRNST.S.Y.N.L.ST-..QN.T..RQ..D.NGSQ-...    LES-1

       210       220       230       240       250       260       270       280       290        300
RTTASGHASTAVGAMSYAQGHFSNAFGTRATAKSAYSLAVGLAATAEGQSTIAIGSDATSSSLGAIALGAGTRAQLQGSIALGQGSVVTQSDNN-----SRPAYT    4223
..R.Q......................Y...EA..........Q.TK..S..V..N.KANAFA.T.I.GN.VVN.GRGV...F..QILDR...-----TDAS..V    Q8
..A.E..........A..K...A......S..EGN.......T.K..KGY......N.QAINY..L....D..VD.DYG....Y..QILNNN..-----NNNK..V    LES-1

       310       320            330       340       350       360        370       380       390       400
PNTQALDPKFQ--ATNNTKAGPL-SIG-SN----SIKRKIINVGAGVNKTDAVNVAQLEAVVKWAKERRITFQGD-DN-STDVKIGLDNTLTIKGGAETNA--LTDNN-IGVV    4223
.LGKT.ADQYK--..RQGDSTDIF...N..NNNS..R.........SRD.........KL.EEL.-N.K...K..G..N.NS.ER..G.......D.Q...---..EA.-....    Q8
.EGNGSNIKSS-K..-----GNG.F...------SST...........YED.........K..ENL..-.Q...K..-..-G.G..KK.GE........-..Q.DK.....N....    LES-1

       410       420       430       440       450       460       470       480       490       500
KEADNSGLKVKLAKTLNNLTEVNTTTLNATTTVKVGSSSSTTAELLSDSLTFTQPNTGSQSTSKTVYGVNGVKFTNNAETTAAIGTTRITRDKIGFARDG    4223
TDGN--.........E.TG..S.-----S..NKIT.SNTNNNN...Q.GG...S-.I..TK-.D....SID.L....DSNSI.TK......KK.....GTN    Q8
TDNN-T........N.SG.ET.S.KN.T.SEK.T...GNN-....Q.GG....-.T.NA-..D.....TD.L...D.SN.ALED-.....K.....SNKA    LES-1

                                                                                            510       520       530
D---------------------------------------------------------------------------------VDEKQAPYLDKKQLKVGSVAITIDNGIDA    4223
.GVDESKPYLDNEKLKVGNSTLNSGSLTVNNT--TGNKQIQVGANGIKFATVANNVANTSATVGTARITEEKIGFAGTNDG...-.......ER....R.E..T.S..N.    Q8
GTVDENKPYLDKDKLKVGNSTLNNGGLTVNNTIGGSNKQIQVGADGIKFADVNVNVSN-AAKFGTTRITEEEIGFADADGK..-.KS........Q..G.K..K.S..N.    LES-1

                        540       550       560       570       580       590       600
GNKKIS-----------NLAKGSSANDAVTIEQLKAAKPTLNAGAGISVTPTEISVDAKSGNVTAPTYNIGVKTTELNSDG    4223
..H..T-----------G.TN.IANT.....K...D........D...INSNNGDLVDS...I.T.....S....K...N.    Q8
.DQ...------------.VKDATDDT....YK...------------------------------------------    LES-1

        610       620       630       640       650       660       670       680       690       700
TSD--KFSVKGSGTNNSLVTAEHLASYLNEVNRTADS-ALQSFTVKEEDDDDANAITVAKDTTKNAGAVSILKLKGKNGLTVATKKD-GTVTFGLSQDSGLTIG    4223
```

```
..GNN....SNAHD.......KD..D...K..E....-..P..K.QNG-.NSN.....G...--.GKTFNT.....E..VNIT.NRAT......ID.SN...TP           Q8
----------------------------Q.QQD..G-.....SIRD.-KGQEFT.SNLYSNGNTPNTFETITFA.E..ISISNDIAK.K.KV.IDPIN...TP           LES-1

              710       720       730       740       750       760       770       780       790       800
KST---------LNNDGLTVKDTNEQIQV--GANGIKFTNVNGSNPGTGIANTARITRDKIGFAGSDGAVDTNKPYLDQDKLQVGNVKITNTGINAGGKAITGLSPTLPSI   4223
.L.VGSDTN-------.NR-LV-I..VP-SADG.ST.NIIK------------------------------------------------------------.........   Q8
.L.VGSDKD-------.K.QLV-I..VASG---.DT.NIIR------------------------------------------------------------.........   LES-1

       810       820       830       840       850       860       870       880       890       900
ADQSS-RNIELGNTIQ-DKDKSNAASINDILNTGFNLKNNNNPIDFVSTYDIVDFANGNATTATVTHDTANKTSKVVYDVNVDDTTIHLTGTDDNKK---LGVKT   4223
.SP.G-...A.....E-E.........D.V..A.......GKDK.......T...ID.........Y.E..Q....A......EK..E...--..G.KQ-.....   Q8
TNAGGV.TT.Q....TS.E...K....G............S.SVG.....NT...ID......K..Y.ET.Q....T......EK..E...--..G.TNKI....   LES-1

       910       920       930       940       950       960       970       980       990      1000
TKLNKTSANGNTATNFNVNSSDED-ALVNAKDIAENLNTLAKEIHTTKGTADTALQTFTVKKVDENNNADDANAITVGQKNANNQ--VNTLTLKGENGLNIKT   4223
I..TE..T...-..T.---.T.D.H...K.S...G......E.........N..............DK...T......KDGTSGK--.....K...K...D...   Q8
.T.TT.N...K-....--STT.N.-...............................K...----DG.T.DET....KDGTQ.GKT....K.......TVA.   LES-1

      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
DKNGTVTFGINTTSGLKAGKST-LNDGGLSIKNPTGSEQIQVGADGVKFAKVNNN---GVVGAGIDGTTRITRDEIGFTGTNGSLDKSKPHLSKDGINAGGKKI------------   4223
..D.........Q......D..T..NN......TASN.............M...----..........................................................------------   Q8
N.D.........Q......D..T..KD.......ASN...............DK----.NSST.....S...K.Q.....A.....TT....T..KLKV.EVE.TNTGINAGGKKI   LES-1

      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
TNIQSGEIAQNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNNFTVSNPYSSYDTSKTSDVITFAGENGITTKVNKGVVRVGIDQTKG   4223
.........K..........................................................................................   Q8
......D.T...N......RV........S..N.A..................H...............................................   LES-1

      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
LTTPKLTVGNNNGKGIVIDSQNGQNTITGLSNTLANVTNDKGSVRTTEQGNIIKDEDKTRAASIVDVLSAGFNLQGNGEAVDFVSTYDTVNFADGNATTA   4223
.................N..................................................................N..T...   Q8
...................KD.................-.AGHALS..LAN-.T........G...N...................D.I.......   LES-1
```

```
       1310      1320       1330      1340      1350      1360      1370      1380      1390      1400
KVTYDDTSKTSKVVYDVNVDDTTIEVK-DKKLGVKTTTLTSTGTGANKFALSNQATGDALVKASDIVAHLNTLSGDIQTAKGASQANNSAGYVDADGNKVI    4223
...........................-......................................................................    Q8
....................NK....TS............K.SANG.ATKF.A-.D.........AT.....A...........SS..S..........    LES-1

       1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
YDSTDNKYYQAKNDGTVDKTKEVAKDKLVAQAQTPDGTLAQMNVKSVINKEQVNDANKKQGINEDNAFVKGLEKAASDNKTKNAAVTVGDLNAVAQTPLT    4223
....................................................................................................    Q8
.....K....VNDK.Q...N...................................I....N..K.T....................    LES-1

       1510      1520      1530      1540      1550      1560      1570      1580      1590
FAGDTGTTAKK-LGETLTIKGGQTDTNKLTDN-NIGVVAGTDGFTVKLAKDLTNLNSVNAGGTKIDDKGVSFVDSSGQAKANTPVLSANGLDL-----    4223
...........-....................-.....................................E..I....AN..................-----    Q8
...........-....................-...................................R..E..I....AN.................GGKRI    LES-1

----------------------------------------------------------------------------------------------------    4223
----------------------------------------------------------------------------------------------------    Q8
SNIGAAVDDNDAVNFKQFNEVAKTVNNLNNQSNSGASLPFVVTDANGKPINGTDGKPQKAIKGADGKYYHANANGVPVDKDGKPITDADKLANLAAHGKP    LES-1

----------------------------------------------------------------------------------------------------    4223
----------------------------------------------------------------------------------------------------    Q8
LDAGHQVVASLGGNSDAITLTNIKSTLPQIDTPNTGNANAGQAQSLPSLSAAQQSNAASVKDVLNVGFNLQTNHNQVDFVKAYDTVNFVNGTGADITSVR    LES-1

----------------------------------------------------------------------------------------------------    4223
----------------------------------------------------------------------------------------------------    Q8
SADGTMSNITVNTALAATDDDGNVLIKAKDGKFYKADDLMPNGSLKAGKSASDAKTPTGLSLVNPNAGKGSTGDAVALNNLSKAVFKSKDGTTTTTVSSD    LES-1
```

```
                     1600
---------------------GGKVISNVG                                                                          4223
---------------------.........                                                                          Q8
GISIQGKDNSSITLSKDGLNV.........                                                                          LES-1

      1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
KGTKDTDAANVQQLNEVRNLLGLGNAGNDNADGNQVNIADIKKDPNSGSSSNRTVIKAGTVLGGKGNNDTEKLATGGIQVGVDKDGNANGDLSNVWVKTQ    4223
.........................---.................................................V......................    Q8
.............................................................................V......................    LES-1

      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
KDGSKKALLATYNAAGQTNYLTNNPAEAIDRINEQGIRFFHVNDGNQEPVVQGRNGIDSSASGKHSVAIGFQAKADGEAAVAIGRQTQAGNQSIAIGDNA    4223
....................V...............................................................................    Q8
....................................................................................................    LES-1

      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
QATGDQSIAIGTGNVVAGKHSGAIGDPSTVKADNSYSVGNNNQFTDATQTDVFGVGNNITVTESNSVALGSNSAISAGTHAGTQAKKSDGTAGTTTTAGA    4223
....................................................................................................    Q8
................T...........................I.......................................................    LES-1

      1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
TGTVKGFAGQTAVGAVSVGASGAERRIQNVAAGEVSATSTDAVNGSQLYKATQSIANATNELDHRIHQNENKANAGISSAMAMASMPQAYIPGRSMVTGG    4223
....................................................................................................    Q8
.....................................................G..............................................    LES-1

      2010      2020      2030      2040
IATHNGQGAVAVGLSKLSDNGQWVFKINGSADTQGHVGAAVGAGFHF*                                                        4223
...............................................*                                                        Q8
...............................................*                                                        LES-1
```

FIGURE 7

Construction of Plasmids Expressing Portions of the 200 kDa Protein Gene from Strain 4223

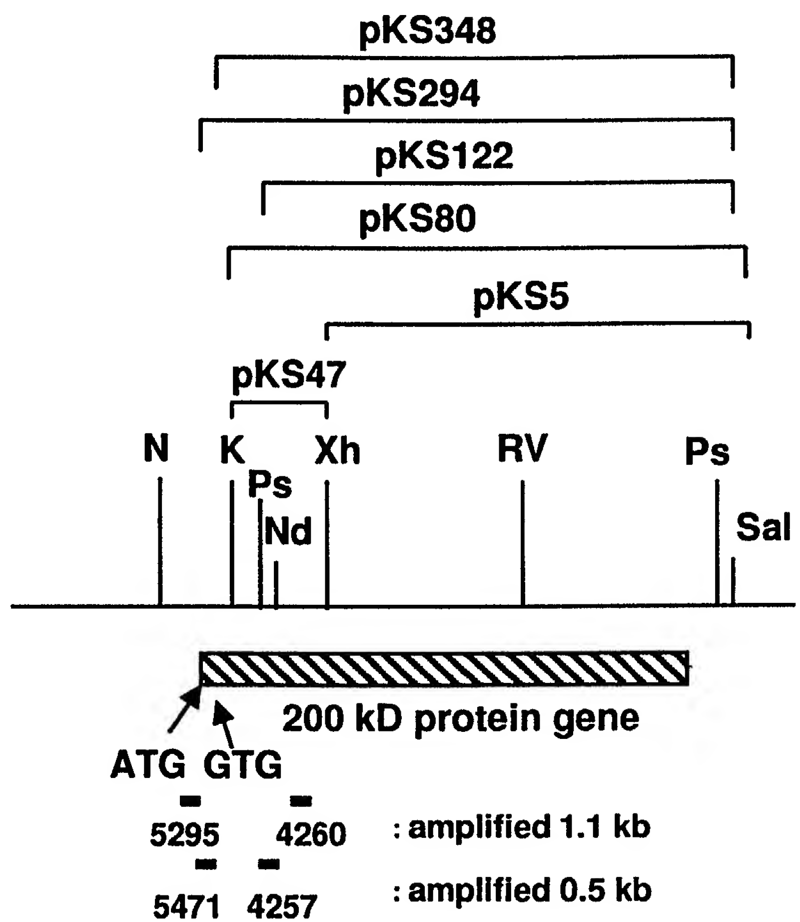


Figure 8. *M. catarrhalis* M56 200kDa gene in pKS348.

```
ATG atc ggt gca acg ctc agt ggc agt gct tat gct caa aaa aaa gat    48
Met Ile Gly Ala Thr Leu Ser Gly Ser Ala Tyr Ala Gln Lys Lys Asp
  1               5                  10                  15

acc aaa cat atc gca att ggt gaa caa aac cag cca aga cgc tca ggc    96
Thr Lys His Ile Ala Ile Gly Glu Gln Asn Gln Pro Arg Arg Ser Gly
             20                  25                  30

act gcc aag gcg gac ggt gat cga gcc att gct att ggt gaa aat gct   144
Thr Ala Lys Ala Asp Gly Asp Arg Ala Ile Ala Ile Gly Glu Asn Ala
         35                  40                  45

aac gca cag ggc ggt caa gcc atc gcc atc ggt agt agt aat aaa act   192
Asn Ala Gln Gly Gly Gln Ala Ile Ala Ile Gly Ser Ser Asn Lys Thr
     50                  55                  60

gtc aat gga agc agt ttg gat aag ata ggt acc gat gct acg ggt caa   240
Val Asn Gly Ser Ser Leu Asp Lys Ile Gly Thr Asp Ala Thr Gly Gln
 65                  70                  75                  80

gag tcc atc gcc atc ggt ggt gat gta aag gct agt ggt gat gcc tcg   288
Glu Ser Ile Ala Ile Gly Gly Asp Val Lys Ala Ser Gly Asp Ala Ser
                 85                  90                  95

att gcc atc ggt agt gat gac tta cat ttg ctt gat cag cat ggt aat   336
Ile Ala Ile Gly Ser Asp Asp Leu His Leu Leu Asp Gln His Gly Asn
            100                 105                 110

cct aaa cat ccg aaa ggt act ctg att aac gat ctt att aac ggc cat   384
Pro Lys His Pro Lys Gly Thr Leu Ile Asn Asp Leu Ile Asn Gly His
        115                 120                 125

gca gta tta aaa gaa ata cga agc tca aag gat aat gat gta aaa tat   432
Ala Val Leu Lys Glu Ile Arg Ser Ser Lys Asp Asn Asp Val Lys Tyr
    130                 135                 140

aga cgc aca acc gca agc gga cac gcc agt act gca gtg gga gcc atg   480
Arg Arg Thr Thr Ala Ser Gly His Ala Ser Thr Ala Val Gly Ala Met
145                 150                 155                 160

tca tat gca cag ggt cat ttt tcc aac gcc ttt ggt aca cgg gca aca   528
Ser Tyr Ala Gln Gly His Phe Ser Asn Ala Phe Gly Thr Arg Ala Thr
                165                 170                 175

gct aaa agt gcc tat tcc ttg gca gtg ggt ctt gcc gcc aca gcc gag   576
Ala Lys Ser Ala Tyr Ser Leu Ala Val Gly Leu Ala Ala Thr Ala Glu
            180                 185                 190

ggc caa tct aca atc gct att ggt tct gat gca aca tct agc tcg ttg   624
Gly Gln Ser Thr Ile Ala Ile Gly Ser Asp Ala Thr Ser Ser Ser Leu
        195                 200                 205

gga gcg ata gcc ctt ggt gca ggt act cgt gct cag cta cag ggc agt   672
Gly Ala Ile Ala Leu Gly Ala Gly Thr Arg Ala Gln Leu Gln Gly Ser
```

```
    210                 215                 220

att gcc cta ggt caa ggt tct gtt gtc act cag agt gat aat aat tct    720
Ile Ala Leu Gly Gln Gly Ser Val Val Thr Gln Ser Asp Asn Asn Ser
225                 230                 235                 240

aga ccg gcc tat aca cca aat acc cag gca cta gac ccc aag ttt caa    768
Arg Pro Ala Tyr Thr Pro Asn Thr Gln Ala Leu Asp Pro Lys Phe Gln
                245                 250                 255

gcc acc aat aat acg aag gcg ggt cca ctt tcc att ggt agt aac tct    816
Ala Thr Asn Asn Thr Lys Ala Gly Pro Leu Ser Ile Gly Ser Asn Ser
            260                 265                 270

atc aaa cgt aaa atc atc aat gtc ggt gca ggt gtt aat aaa acc gat    864
Ile Lys Arg Lys Ile Ile Asn Val Gly Ala Gly Val Asn Lys Thr Asp
        275                 280                 285

gcg gtc aat gtg gca cag cta gaa gcg gtg gtg aag tgg gct aag gag    912
Ala Val Asn Val Ala Gln Leu Glu Ala Val Val Lys Trp Ala Lys Glu
    290                 295                 300

cgt aga att act ttt cag ggt gat gat aac agt act gac gta aaa ata    960
Arg Arg Ile Thr Phe Gln Gly Asp Asp Asn Ser Thr Asp Val Lys Ile
305                 310                 315                 320

ggt ttg gat aat act tta act att aaa ggt ggt gca gag acc aac gca    1008
Gly Leu Asp Asn Thr Leu Thr Ile Lys Gly Gly Ala Glu Thr Asn Ala
                325                 330                 335

tta acc gat aat aat atc ggt gtg gta aaa gag gct gat aat agt ggt    1056
Leu Thr Asp Asn Asn Ile Gly Val Val Lys Glu Ala Asp Asn Ser Gly
            340                 345                 350

ctg aaa gtt aaa ctt gct aaa act tta aac aat ctt act gag gtg aat    1104
Leu Lys Val Lys Leu Ala Lys Thr Leu Asn Asn Leu Thr Glu Val Asn
        355                 360                 365

aca act aca tta aat gcc aca acc aca gtt aag gta ggt agt agt agt    1152
Thr Thr Thr Leu Asn Ala Thr Thr Thr Val Lys Val Gly Ser Ser Ser
    370                 375                 380

agt act aca gct gaa tta ttg agt gat agt tta acc ttt acc cag ccc    1200
Ser Thr Thr Ala Glu Leu Leu Ser Asp Ser Leu Thr Phe Thr Gln Pro
385                 390                 395                 400

aat aca ggc agt caa agc aca agc aaa acc gtc tat ggc gtt aat ggg    1248
Asn Thr Gly Ser Gln Ser Thr Ser Lys Thr Val Tyr Gly Val Asn Gly
                405                 410                 415

gtg aag ttt act aat aat gca gaa aca aca gca gca atc ggc act act    1296
Val Lys Phe Thr Asn Asn Ala Glu Thr Thr Ala Ala Ile Gly Thr Thr
            420                 425                 430

cgt att acc aga gat aaa att ggc ttt gct cga gat ggt gat gtt gat    1344
Arg Ile Thr Arg Asp Lys Ile Gly Phe Ala Arg Asp Gly Asp Val Asp
        435                 440                 445
```

```
gaa aaa caa gca cca tat ttg gat aaa aaa caa ctt aaa gtg ggt agt    1392
Glu Lys Gln Ala Pro Tyr Leu Asp Lys Lys Gln Leu Lys Val Gly Ser
    450                 455                 460

gtt gca att acc ata gac aat ggc att gat gca ggt aat aaa aag atc    1440
Val Ala Ile Thr Ile Asp Asn Gly Ile Asp Ala Gly Asn Lys Lys Ile
465                 470                 475                 480

agt aat ctt gcc aaa ggt agc agt gct aac gat gcg gtt acc atc gaa    1488
Ser Asn Leu Ala Lys Gly Ser Ser Ala Asn Asp Ala Val Thr Ile Glu
                485                 490                 495

cag ctc aaa gcc gcc aag cct act tta aac gca ggc gct ggc atc agt    1536
Gln Leu Lys Ala Ala Lys Pro Thr Leu Asn Ala Gly Ala Gly Ile Ser
            500                 505                 510

gtc aca cct act gaa ata tca gtt gat gct aag agt ggc aat gtt acc    1584
Val Thr Pro Thr Glu Ile Ser Val Asp Ala Lys Ser Gly Asn Val Thr
        515                 520                 525

gcc cca act tac aac att ggc gtg aaa acc acc gag ctt aac agt gat    1632
Ala Pro Thr Tyr Asn Ile Gly Val Lys Thr Thr Glu Leu Asn Ser Asp
    530                 535                 540

ggc act agt gat aaa ttt agt gtt aag ggt agt ggt acg aac aat agc    1680
Gly Thr Ser Asp Lys Phe Ser Val Lys Gly Ser Gly Thr Asn Asn Ser
545                 550                 555                 560

tta gtt acc gcc gaa cat ttg gca agc tat cta aat gaa gtc aat cga    1728
Leu Val Thr Ala Glu His Leu Ala Ser Tyr Leu Asn Glu Val Asn Arg
                565                 570                 575

acg gct gac agt gct cta caa agc ttt acc gtt aaa gaa gaa gac gat    1776
Thr Ala Asp Ser Ala Leu Gln Ser Phe Thr Val Lys Glu Glu Asp Asp
            580                 585                 590

gat gac gcc aac gct atc acc gtg gct aaa gat acg aca aaa aat gcc    1824
Asp Asp Ala Asn Ala Ile Thr Val Ala Lys Asp Thr Thr Lys Asn Ala
        595                 600                 605

ggc gca gtc agc atc tta aaa ctc aaa ggt aaa aac ggt cta acg gtt    1872
Gly Ala Val Ser Ile Leu Lys Leu Lys Gly Lys Asn Gly Leu Thr Val
    610                 615                 620

gct acc aaa aaa gat ggt acg gtt acc ttt ggg ctt agc caa gat agc    1920
Ala Thr Lys Lys Asp Gly Thr Val Thr Phe Gly Leu Ser Gln Asp Ser
625                 630                 635                 640

ggt ctg acc att ggc aaa agc acc cta aac aac gat ggc ttg act gtt    1968
Gly Leu Thr Ile Gly Lys Ser Thr Leu Asn Asn Asp Gly Leu Thr Val
                645                 650                 655

aaa gat acc aac gaa caa atc caa gtc ggt gct aat ggc att aaa ttt    2016
Lys Asp Thr Asn Glu Gln Ile Gln Val Gly Ala Asn Gly Ile Lys Phe
            660                 665                 670
```

act aat gtg aat ggt agt aat cca ggt act ggc att gca aat acc gct 2064
Thr Asn Val Asn Gly Ser Asn Pro Gly Thr Gly Ile Ala Asn Thr Ala
675 680 685

cgc att acc aga gat aaa att ggc ttt gct ggt tct gat ggt gca gtt 2112
Arg Ile Thr Arg Asp Lys Ile Gly Phe Ala Gly Ser Asp Gly Ala Val
690 695 700

gat aca aac aaa cct tat ctt gat caa gac aag cta caa gtt ggc aat 2160
Asp Thr Asn Lys Pro Tyr Leu Asp Gln Asp Lys Leu Gln Val Gly Asn
705 710 715 720

gtt aag att acc aac act ggc att aac gca ggt ggt aaa gcc atc aca 2208
Val Lys Ile Thr Asn Thr Gly Ile Asn Ala Gly Gly Lys Ala Ile Thr
725 730 735

ggg ctg tcc cca aca ctg cct agc att gcc gat caa agt agc cgc aac 2256
Gly Leu Ser Pro Thr Leu Pro Ser Ile Ala Asp Gln Ser Ser Arg Asn
740 745 750

ata gaa ctg ggc aat aca atc caa gac aaa gac aaa tcc aac gct gcc 2304
Ile Glu Leu Gly Asn Thr Ile Gln Asp Lys Asp Lys Ser Asn Ala Ala
755 760 765

agc att aat gat ata tta aat aca ggc ttt aac cta aaa aat aat aac 2352
Ser Ile Asn Asp Ile Leu Asn Thr Gly Phe Asn Leu Lys Asn Asn Asn
770 775 780

aac ccc att gac ttt gtc tcc act tat gac att gtt gac ttt gcc aat 2400
Asn Pro Ile Asp Phe Val Ser Thr Tyr Asp Ile Val Asp Phe Ala Asn
785 790 795 800

ggc aat gcc acc acc gcc aca gta acc cat gat acc gct aac aaa acc 2448
Gly Asn Ala Thr Thr Ala Thr Val Thr His Asp Thr Ala Asn Lys Thr
805 810 815

agt aaa gtg gta tat gat gtg aat gtg gat gat aca acc att cat cta 2496
Ser Lys Val Val Tyr Asp Val Asn Val Asp Asp Thr Thr Ile His Leu
820 825 830

aca ggc act gat gac aat aaa aaa ctt ggc gtc aaa acc acc aaa ctg 2544
Thr Gly Thr Asp Asp Asn Lys Lys Leu Gly Val Lys Thr Thr Lys Leu
835 840 845

aac aaa aca agt gct aat ggt aat aca gca act aac ttt aat gtt aac 2592
Asn Lys Thr Ser Ala Asn Gly Asn Thr Ala Thr Asn Phe Asn Val Asn
850 855 860

tct agt gat gaa gat gcc ctt gtt aac gcc aaa gac atc gcc gaa aat 2640
Ser Ser Asp Glu Asp Ala Leu Val Asn Ala Lys Asp Ile Ala Glu Asn
865 870 875 880

cta aac acc cta gcc aag gaa att cac acc acc aaa ggc aca gca gac 2688
Leu Asn Thr Leu Ala Lys Glu Ile His Thr Thr Lys Gly Thr Ala Asp
885 890 895

acc gcc cta caa acc ttt acc gtt aaa aag gta gat gaa aat aat aat 2736

Thr Ala Leu Gln Thr Phe Thr Val Lys Lys Val Asp Glu Asn Asn Asn
            900   905   910

gct gat gac gcc aac gcc atc acc gtg ggt caa aag aac gca aat aat   2784
Ala Asp Asp Ala Asn Ala Ile Thr Val Gly Gln Lys Asn Ala Asn Asn
        915   920   925

caa gtc aac acc cta aca ctc aaa ggt gaa aac ggt ctt aat att aaa   2832
Gln Val Asn Thr Leu Thr Leu Lys Gly Glu Asn Gly Leu Asn Ile Lys
    930   935   940

acc gac aaa aat ggt acg gtt acc ttt ggc att aac acc aca agc ggt   2880
Thr Asp Lys Asn Gly Thr Val Thr Phe Gly Ile Asn Thr Thr Ser Gly
945   950   955   960

ctt aaa gcc ggc aaa agc acc cta aac gac ggt ggc ttg tct att aaa   2928
Leu Lys Ala Gly Lys Ser Thr Leu Asn Asp Gly Gly Leu Ser Ile Lys
                965   970   975

aac ccc act ggt agc gaa caa atc caa gtc ggt gct gat ggc gtg aag   2976
Asn Pro Thr Gly Ser Glu Gln Ile Gln Val Gly Ala Asp Gly Val Lys
            980   985   990

ttt gcc aag gtt aat aat aat ggt gtt gta ggt gct ggc att gat ggc   3024
Phe Ala Lys Val Asn Asn Asn Gly Val Val Gly Ala Gly Ile Asp Gly
        995   1000   1005

aca act cgc att acc aga gat gaa att ggc ttt act ggg act aat ggc   3072
Thr Thr Arg Ile Thr Arg Asp Glu Ile Gly Phe Thr Gly Thr Asn Gly
    1010   1015   1020

tca ctt gat aaa agc aaa ccc cac cta agc aaa gac ggc att aac gca   3120
Ser Leu Asp Lys Ser Lys Pro His Leu Ser Lys Asp Gly Ile Asn Ala
1025   1030   1035   1040

ggt ggt aaa aag att acc aac att caa tca ggt gag att gcc caa aac   3168
Gly Gly Lys Lys Ile Thr Asn Ile Gln Ser Gly Glu Ile Ala Gln Asn
                1045   1050   1055

agc cat gat gct gtg aca ggc ggc aag att tat gat tta aaa acc gaa   3216
Ser His Asp Ala Val Thr Gly Gly Lys Ile Tyr Asp Leu Lys Thr Glu
            1060   1065   1070

ctt gaa aac aaa atc agc agt act gcc aaa aca gca caa aac tca tta   3264
Leu Glu Asn Lys Ile Ser Ser Thr Ala Lys Thr Ala Gln Asn Ser Leu
        1075   1080   1085

cac gaa ttc tca gta gca gat gaa caa ggt aat aac ttt acg gtt agt   3312
His Glu Phe Ser Val Ala Asp Glu Gln Gly Asn Asn Phe Thr Val Ser
    1090   1095   1100

aac cct tac tcc agt tat gac acc tca aag acc tct gat gtc atc acc   3360
Asn Pro Tyr Ser Ser Tyr Asp Thr Ser Lys Thr Ser Asp Val Ile Thr
1105   1110   1115   1120

ttt gca ggt gaa aac ggc att acc acc aag gta aat aaa ggt gtg gtg   3408
Phe Ala Gly Glu Asn Gly Ile Thr Thr Lys Val Asn Lys Gly Val Val

1125 1130 1135

cgt gtg ggc att gac caa acc aaa ggc tta acc acg cct aag ctg acc 3456
Arg Val Gly Ile Asp Gln Thr Lys Gly Leu Thr Thr Pro Lys Leu Thr
1140 1145 1150

gtg ggt aat aat aat ggc aaa ggc att gtc att gac agc caa aat ggt 3504
Val Gly Asn Asn Asn Gly Lys Gly Ile Val Ile Asp Ser Gln Asn Gly
1155 1160 1165

caa aat acc atc aca gga cta agc aac act cta gct aat gtt acc aat 3552
Gln Asn Thr Ile Thr Gly Leu Ser Asn Thr Leu Ala Asn Val Thr Asn
1170 1175 1180

gat aaa ggt agc gta cgc acc aca gaa cag ggc aat ata atc aaa gac 3600
Asp Lys Gly Ser Val Arg Thr Thr Glu Gln Gly Asn Ile Ile Lys Asp
1185 1190 1195 1200

gaa gac aaa acc cgt gcc gcc agc att gtt gat gtg cta agc gca ggc 3648
Glu Asp Lys Thr Arg Ala Ala Ser Ile Val Asp Val Leu Ser Ala Gly
1205 1210 1215

ttt aac ttg caa ggc aat ggt gaa gcg gtt gac ttt gtc tcc act tat 3696
Phe Asn Leu Gln Gly Asn Gly Glu Ala Val Asp Phe Val Ser Thr Tyr
1220 1225 1230

gac acc gtc aac ttt gcc gat ggc aat gcc acc acc gct aag gtg acc 3744
Asp Thr Val Asn Phe Ala Asp Gly Asn Ala Thr Thr Ala Lys Val Thr
1235 1240 1245

tat gat gac aca agc aaa acc agt aaa gtg gtc tat gat gtc aat gtg 3792
Tyr Asp Asp Thr Ser Lys Thr Ser Lys Val Val Tyr Asp Val Asn Val
1250 1255 1260

gat gat aca acc att gaa gtt aaa gat aaa aaa ctt ggc gta aaa acc 3840
Asp Asp Thr Thr Ile Glu Val Lys Asp Lys Lys Leu Gly Val Lys Thr
1265 1270 1275 1280

acc aca ttg acc agt act ggc aca ggt gct aat aaa ttt gcc cta agc 3888
Thr Thr Leu Thr Ser Thr Gly Thr Gly Ala Asn Lys Phe Ala Leu Ser
1285 1290 1295

aat caa gct act ggc gat gcg ctt gtc aag gcc agt gat atc gtt gct 3936
Asn Gln Ala Thr Gly Asp Ala Leu Val Lys Ala Ser Asp Ile Val Ala
1300 1305 1310

cat cta aac acc tta tct ggc gac atc caa act gcc aaa ggg gca agc 3984
His Leu Asn Thr Leu Ser Gly Asp Ile Gln Thr Ala Lys Gly Ala Ser
1315 1320 1325

caa gcg aac aac tca gca ggc tat gtg gat gct gat ggc aat aag gtc 4032
Gln Ala Asn Asn Ser Ala Gly Tyr Val Asp Ala Asp Gly Asn Lys Val
1330 1335 1340

atc tat gac agt acc gat aac aag tac tat caa gcc aaa aat gat ggc 4080
Ile Tyr Asp Ser Thr Asp Asn Lys Tyr Tyr Gln Ala Lys Asn Asp Gly
1345 1350 1355 1360

aca gtt gat aaa acc aaa gaa gtt gcc aaa gac aaa ctg gtc gcc caa 4128
Thr Val Asp Lys Thr Lys Glu Val Ala Lys Asp Lys Leu Val Ala Gln
1365 1370 1375

gcc caa acc cca gat ggc aca ttg gct caa atg aat gtc aaa tca gtc 4176
Ala Gln Thr Pro Asp Gly Thr Leu Ala Gln Met Asn Val Lys Ser Val
1380 1385 1390

att aac aaa gaa caa gta aat gat gcc aat aaa aag caa ggc atc aat 4224
Ile Asn Lys Glu Gln Val Asn Asp Ala Asn Lys Lys Gln Gly Ile Asn
1395 1400 1405

gaa gac aac gcc ttt gtt aaa gga ctt gaa aaa gcc gct tct gat aac 4272
Glu Asp Asn Ala Phe Val Lys Gly Leu Glu Lys Ala Ala Ser Asp Asn
1410 1415 1420

aaa acc aaa aac gcc gca gta act gtg ggt gat tta aat gcc gtt gcc 4320
Lys Thr Lys Asn Ala Ala Val Thr Val Gly Asp Leu Asn Ala Val Ala
1425 1430 1435 1440

caa aca ccg ctg acc ttt gca ggg gat aca ggc aca acg gct aaa aaa 4368
Gln Thr Pro Leu Thr Phe Ala Gly Asp Thr Gly Thr Thr Ala Lys Lys
1445 1450 1455

ctg ggc gag act ttg acc atc aaa ggt ggg caa aca gac acc aat aag 4416
Leu Gly Glu Thr Leu Thr Ile Lys Gly Gly Gln Thr Asp Thr Asn Lys
1460 1465 1470

cta acc gat aat aac atc ggt gtg gta gca ggt act gat ggc ttc act 4464
Leu Thr Asp Asn Asn Ile Gly Val Val Ala Gly Thr Asp Gly Phe Thr
1475 1480 1485

gtc aaa ctt gcc aaa gac cta acc aat ctt aac agc gtt aat gca ggt 4512
Val Lys Leu Ala Lys Asp Leu Thr Asn Leu Asn Ser Val Asn Ala Gly
1490 1495 1500

ggc acc aaa att gat gac aaa ggc gtg tct ttt gta gac tca agc ggt 4560
Gly Thr Lys Ile Asp Asp Lys Gly Val Ser Phe Val Asp Ser Ser Gly
1505 1510 1515 1520

caa gcc aaa gca aac acc cct gtg cta agt gcc aat ggg ctg gac ctg 4608
Gln Ala Lys Ala Asn Thr Pro Val Leu Ser Ala Asn Gly Leu Asp Leu
1525 1530 1535

ggt ggc aag gtc atc agt aat gtg ggc aaa ggc aca aaa gat acc gac 4656
Gly Gly Lys Val Ile Ser Asn Val Gly Lys Gly Thr Lys Asp Thr Asp
1540 1545 1550

gct gcc aat gta caa cag tta aac gaa gta cgc aac ttg ttg ggt ctt 4704
Ala Ala Asn Val Gln Gln Leu Asn Glu Val Arg Asn Leu Leu Gly Leu
1555 1560 1565

ggt aat gct ggt aat gat aac gct gac ggc aat cag gta aac att gcc 4752
Gly Asn Ala Gly Asn Asp Asn Ala Asp Gly Asn Gln Val Asn Ile Ala
1570 1575 1580

gac atc aaa aaa gac cca aat tca ggt tca tca tct aac cgc act gtc 4800
Asp Ile Lys Lys Asp Pro Asn Ser Gly Ser Ser Ser Asn Arg Thr Val
1585 1590 1595 1600

atc aaa gca ggc acg gta ctt ggc ggt aaa ggt aat aac gat acc gaa 4848
Ile Lys Ala Gly Thr Val Leu Gly Gly Lys Gly Asn Asn Asp Thr Glu
1605 1610 1615

aaa ctt gcc act ggt ggt ata caa gtg ggc gtg gat aaa gac ggc aac 4896
Lys Leu Ala Thr Gly Gly Ile Gln Val Gly Val Asp Lys Asp Gly Asn
1620 1625 1630

gct aac ggc gat tta agc aat gtt tgg gtc aaa acc caa aaa gat ggc 4944
Ala Asn Gly Asp Leu Ser Asn Val Trp Val Lys Thr Gln Lys Asp Gly
1635 1640 1645

agc aaa aaa gcc ctg ctc gcc act tat aac gcc gca ggt cag acc aac 4992
Ser Lys Lys Ala Leu Leu Ala Thr Tyr Asn Ala Ala Gly Gln Thr Asn
1650 1655 1660

tat ttg acc aac aac ccc gca gaa gcc att gac aga ata aat gaa caa 5040
Tyr Leu Thr Asn Asn Pro Ala Glu Ala Ile Asp Arg Ile Asn Glu Gln
1665 1670 1675 1680

ggt atc cgc ttc ttc cat gtc aac gat ggc aat caa gag cct gtg gta 5088
Gly Ile Arg Phe Phe His Val Asn Asp Gly Asn Gln Glu Pro Val Val
1685 1690 1695

caa ggg cgt aac ggc att gac tca agt gcc tca ggc aag cac tca gtg 5136
Gln Gly Arg Asn Gly Ile Asp Ser Ser Ala Ser Gly Lys His Ser Val
1700 1705 1710

gcg ata ggt ttc cag gcc aag gca gat ggt gaa gcc gcc gtt gcc ata 5184
Ala Ile Gly Phe Gln Ala Lys Ala Asp Gly Glu Ala Ala Val Ala Ile
1715 1720 1725

ggc aga caa acc caa gca ggc aac caa tcc atc gcc atc ggt gat aac 5232
Gly Arg Gln Thr Gln Ala Gly Asn Gln Ser Ile Ala Ile Gly Asp Asn
1730 1735 1740

gca caa gcc acg ggc gat caa tcc atc gcc atc ggt aca ggc aat gtg 5280
Ala Gln Ala Thr Gly Asp Gln Ser Ile Ala Ile Gly Thr Gly Asn Val
1745 1750 1755 1760

gta gca ggt aag cac tct ggt gcc atc ggc gac cca agc act gtt aag 5328
Val Ala Gly Lys His Ser Gly Ala Ile Gly Asp Pro Ser Thr Val Lys
1765 1770 1775

gct gat aac agt tac agt gtg ggt aat aac aac cag ttt acc gat gcc 5376
Ala Asp Asn Ser Tyr Ser Val Gly Asn Asn Asn Gln Phe Thr Asp Ala
1780 1785 1790

act caa acc gat gtc ttt ggt gtg ggc aat aac atc acc gtg acc gaa 5424
Thr Gln Thr Asp Val Phe Gly Val Gly Asn Asn Ile Thr Val Thr Glu
1795 1800 1805

agt aac tcg gtt gcc tta ggt tca aac tct gcc atc agt gca ggc aca 5472

Ser Asn Ser Val Ala Leu Gly Ser Asn Ser Ala Ile Ser Ala Gly Thr
1810 1815 1820

cac gca ggc aca caa gcc aaa aaa tct gac ggc aca gca ggt aca acc 5520
His Ala Gly Thr Gln Ala Lys Lys Ser Asp Gly Thr Ala Gly Thr Thr
1825 1830 1835 1840

acc aca gca ggt gca acc ggt acg gtt aaa ggc ttt gct gga caa acg 5568
Thr Thr Ala Gly Ala Thr Gly Thr Val Lys Gly Phe Ala Gly Gln Thr
1845 1850 1855

gcg gtt ggt gcg gtc tcc gtg ggt gcc tca ggt gct gaa cgc cgt atc 5616
Ala Val Gly Ala Val Ser Val Gly Ala Ser Gly Ala Glu Arg Arg Ile
1860 1865 1870

caa aat gtg gca gca ggt gag gtc agt gcc acc agc acc gat gcg gtc 5664
Gln Asn Val Ala Ala Gly Glu Val Ser Ala Thr Ser Thr Asp Ala Val
1875 1880 1885

aat ggt agc cag ttg tac aaa gcc acc caa agc att gcc aac gca acc 5712
Asn Gly Ser Gln Leu Tyr Lys Ala Thr Gln Ser Ile Ala Asn Ala Thr
1890 1895 1900

aat gag ctt gac cat cgt atc cac caa aac gaa aat aag gcc aat gca 5760
Asn Glu Leu Asp His Arg Ile His Gln Asn Glu Asn Lys Ala Asn Ala
1905 1910 1915 1920

ggg att tca tca gcg atg gcg atg gcg tcc atg cca caa gcc tac att 5808
Gly Ile Ser Ser Ala Met Ala Met Ala Ser Met Pro Gln Ala Tyr Ile
1925 1930 1935

cct ggc aga tcc atg gtt acc ggg ggt att gcc acc cac aac ggt caa 5856
Pro Gly Arg Ser Met Val Thr Gly Gly Ile Ala Thr His Asn Gly Gln
1940 1945 1950

ggt gcg gtg gca gtg gga ctg tcg aag ctg tcg gat aat ggt caa tgg 5904
Gly Ala Val Ala Val Gly Leu Ser Lys Leu Ser Asp Asn Gly Gln Trp
1955 1960 1965

gta ttt aaa atc aat ggt tca gcc gat acc caa ggc cat gta ggg gcg 5952
Val Phe Lys Ile Asn Gly Ser Ala Asp Thr Gln Gly His Val Gly Ala
1970 1975 1980

gca gtt ggt gca ggt ttt cac ttt taagccataa atcgcaagat tttacttaaa 6006
Ala Val Gly Ala Gly Phe His Phe
1985 1990

aatcaatctc accatagttg tataaaacag catcagcatc agtcatatta ctgatgctga 6066

tgttttttat cacttaaacc attttaccgc tcaagtgatt ctctttcacc atgaccaaat 6126

cgccattgat cataggtaaa cttattgagt aaattttatc aatgtagttg ttagatatgg 6186

ttaaaattgt gccattgacc aaaaaatgac cgatttatcc cgaaaatttc tgattatgat 6246

ccgttgacct gca 6259

Figure 9A Construction of pKS294

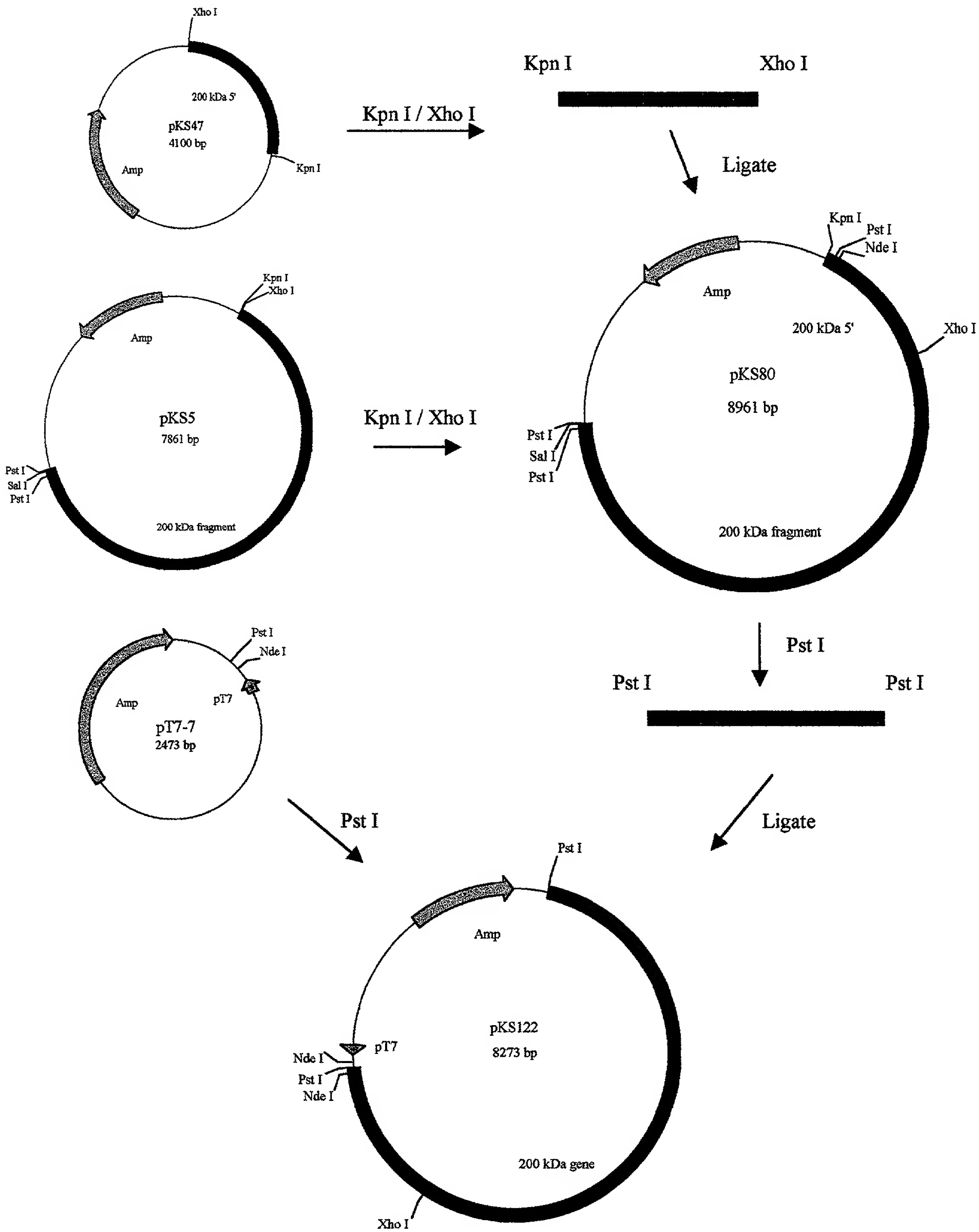


Figure 9B Construction of pKS294

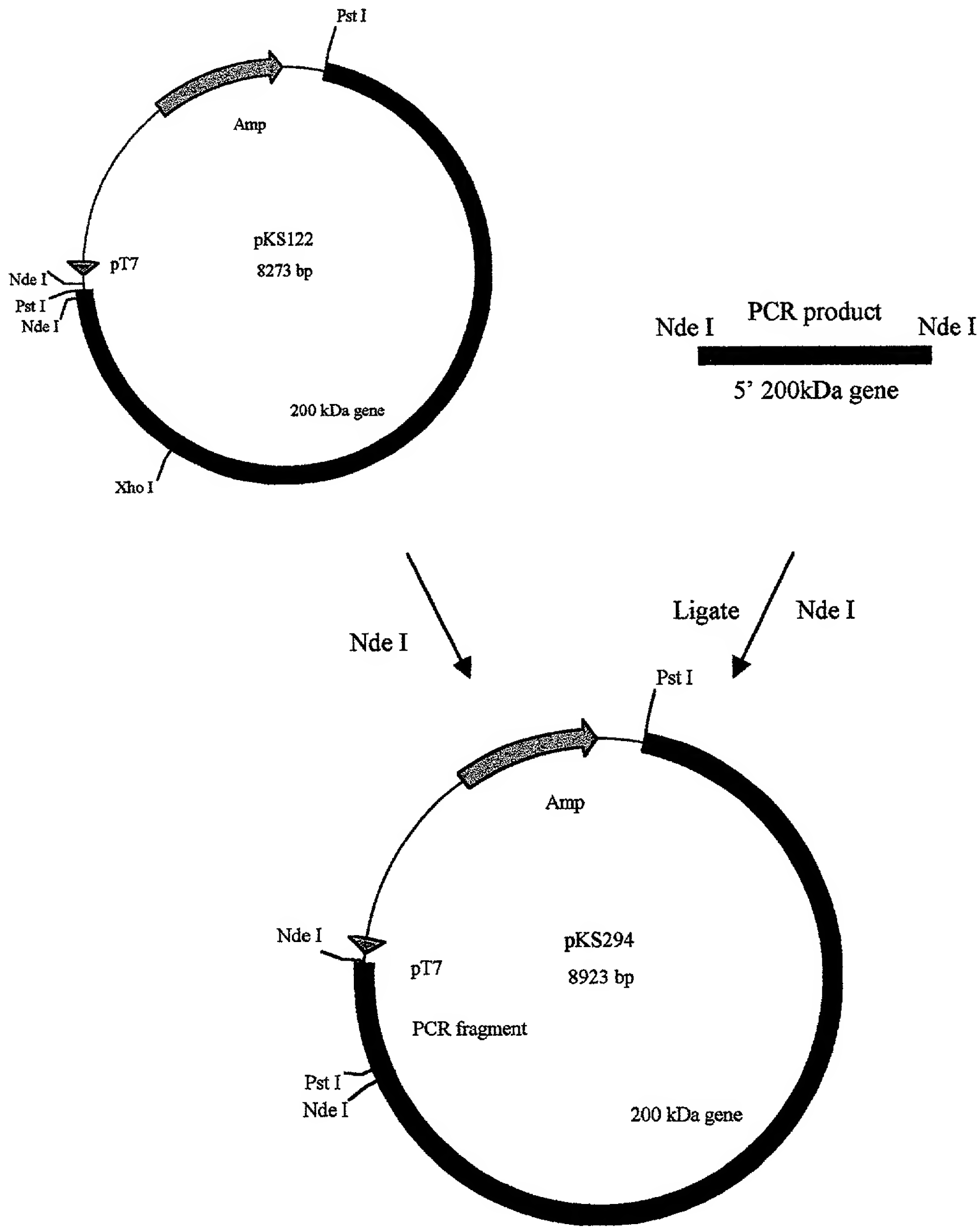


Figure 10. Construction of pKS348

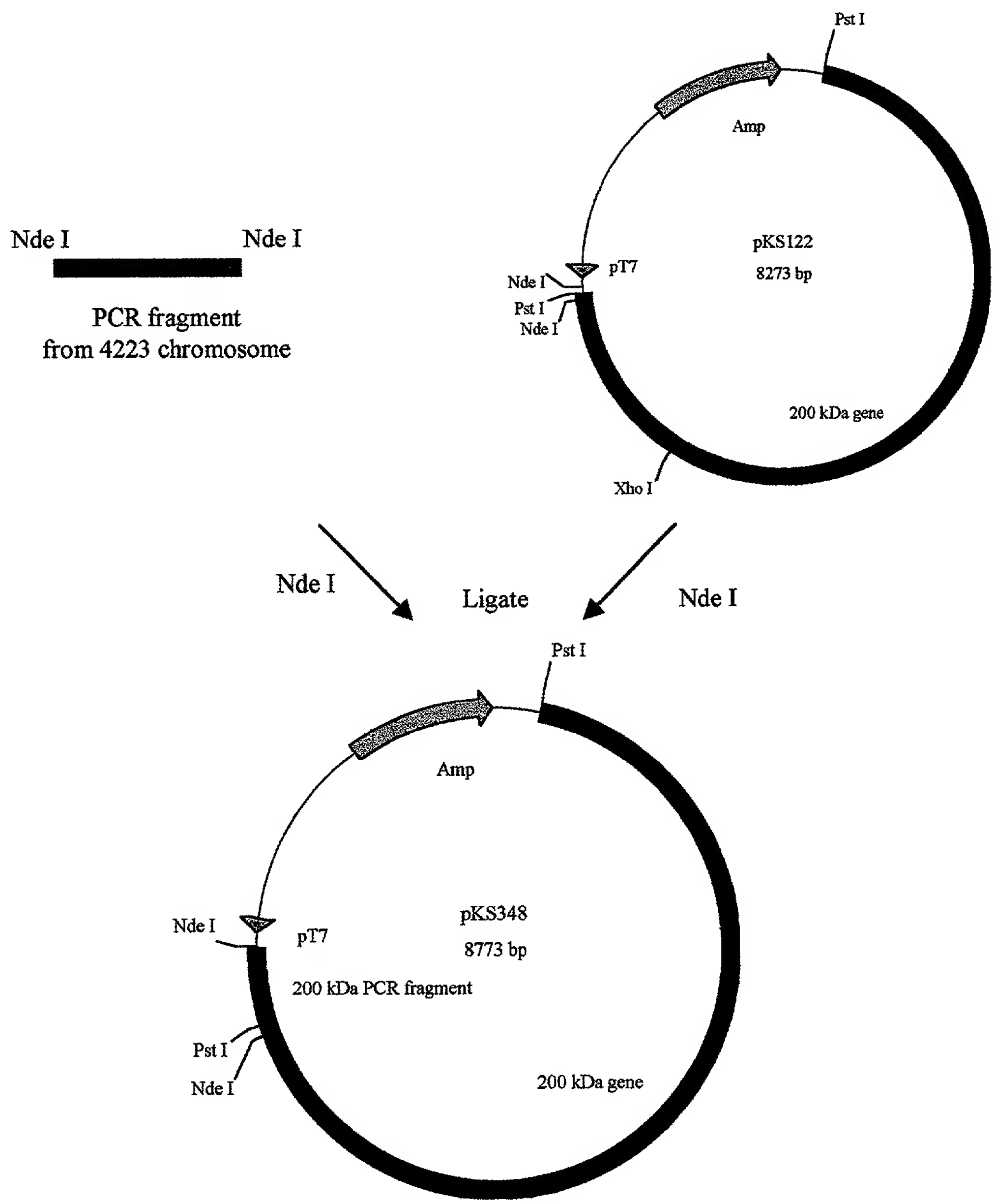


FIGURE 11

Purification of r200 kDa Protein from *E. coli*

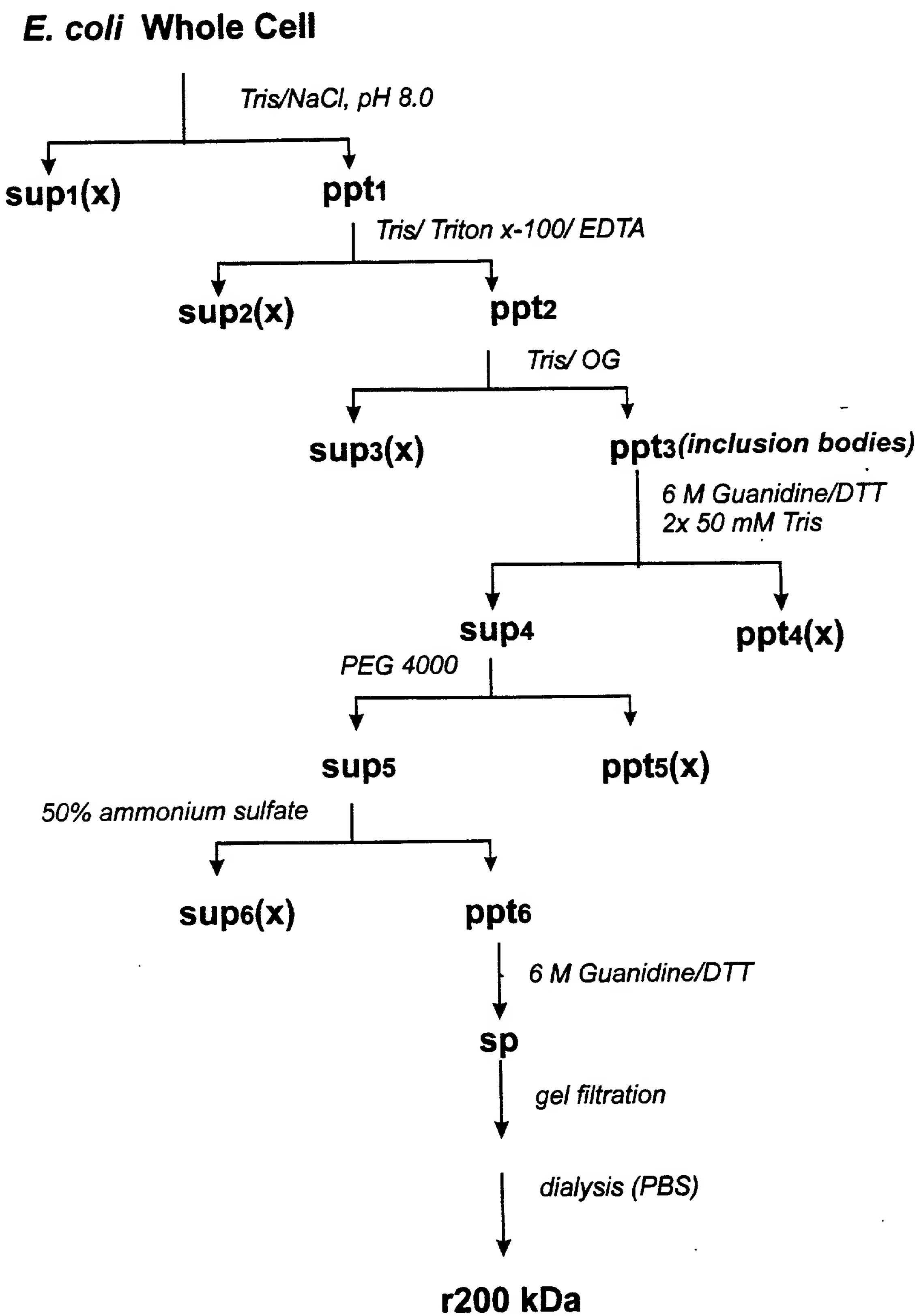
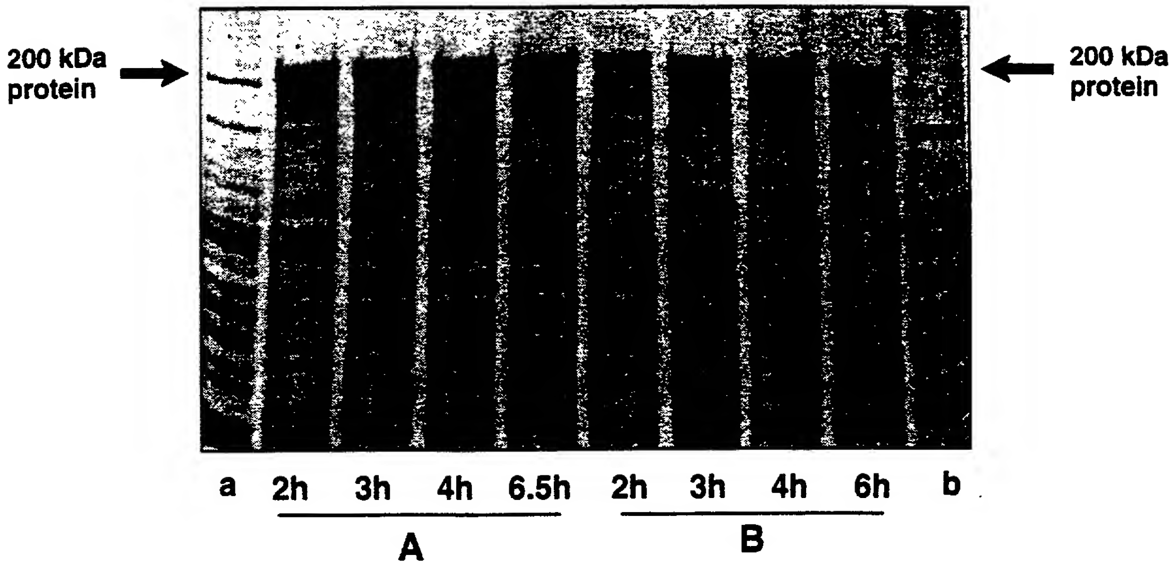


FIGURE 12

Expression of M56 r200 kDa Protein Gene in *E. coli*



A: KS358 induced when O.D. at 600 nm was 0.26

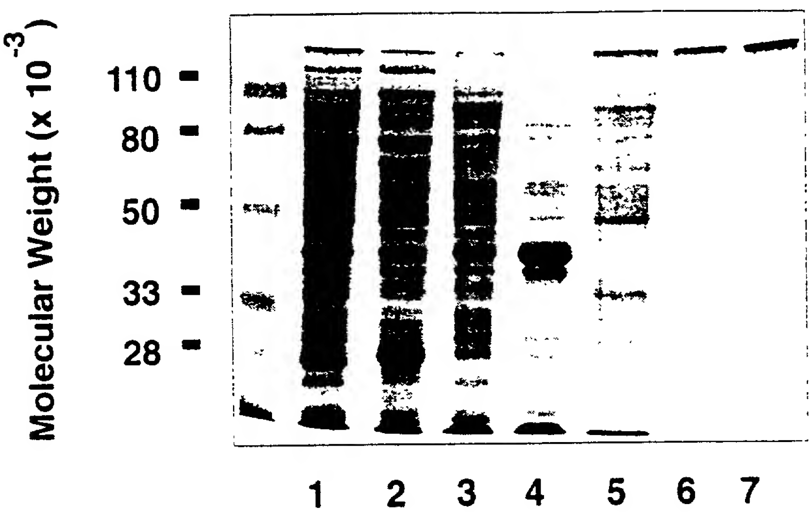
B: KS358 induced when O.D. at 600 nm was 0.44

a: strain 4223 lysate

b: KS358 cultured overnight

FIGURE 13

Purification of M56 r200 kDa Protein (4223)



1. *E. coli* **Whole cells**

2. **Soluble proteins after 50 mM Tris/ NaCl, pH 8, extraction**

3. **Soluble proteins after Tris/ Triton X-100/ EDTA extraction**

4. **Soluble proteins after Tris/ OG extraction**

5. **Pellet after Tris/ OG extraction**

6-7. **Purified 200 kDa protein**

FIGURE 14

Anti-M56 r200 kDa Antibody Titers in Mice

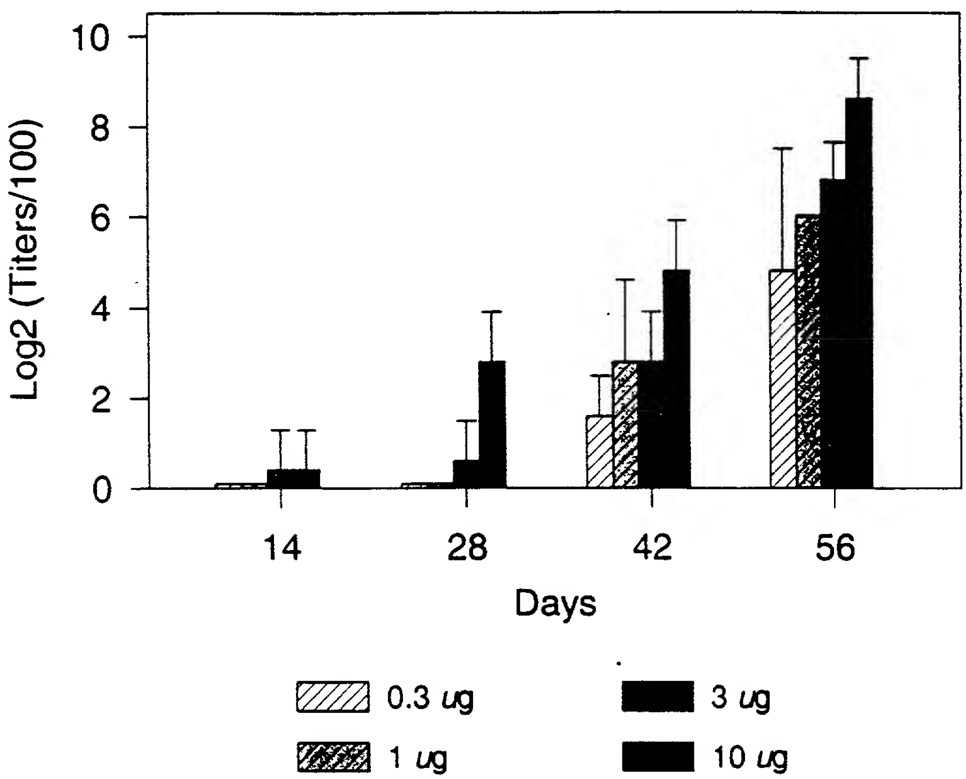


FIGURE 15

Anti-M56 r200 kDa Antibody Titers in Guinea Pigs

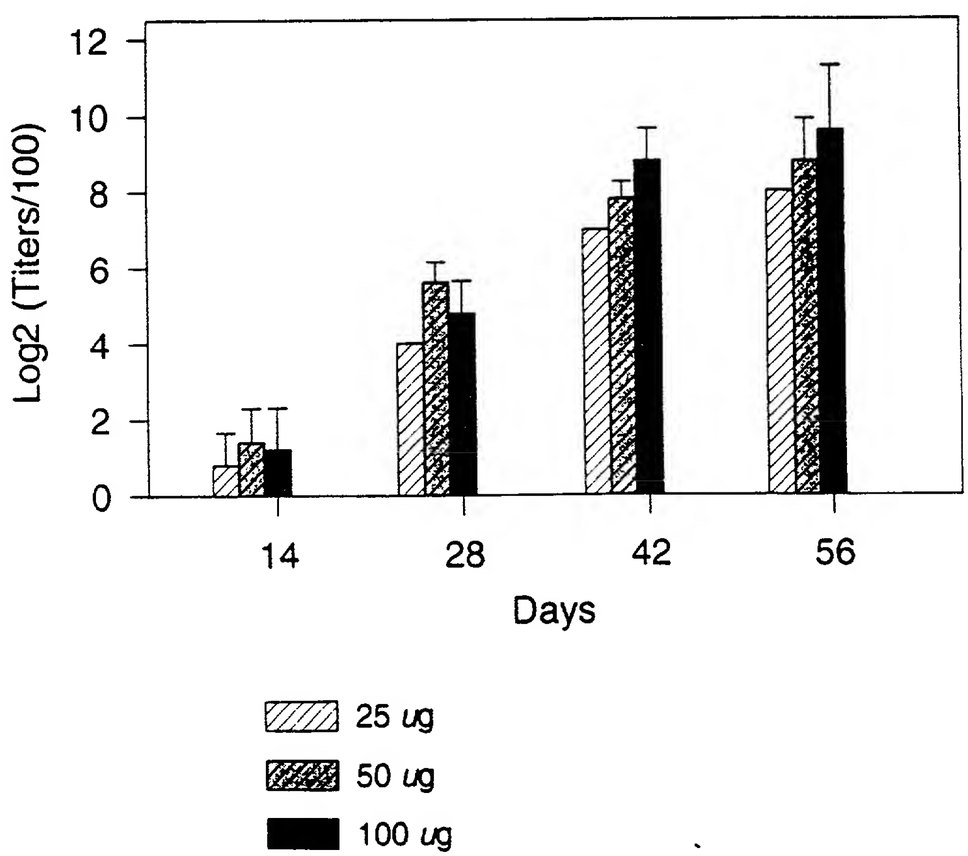


FIGURE 16

PCR amplification of DNA fragments carrying a portion of the 200 kDa protein gene from chromosomal DNA of RH408

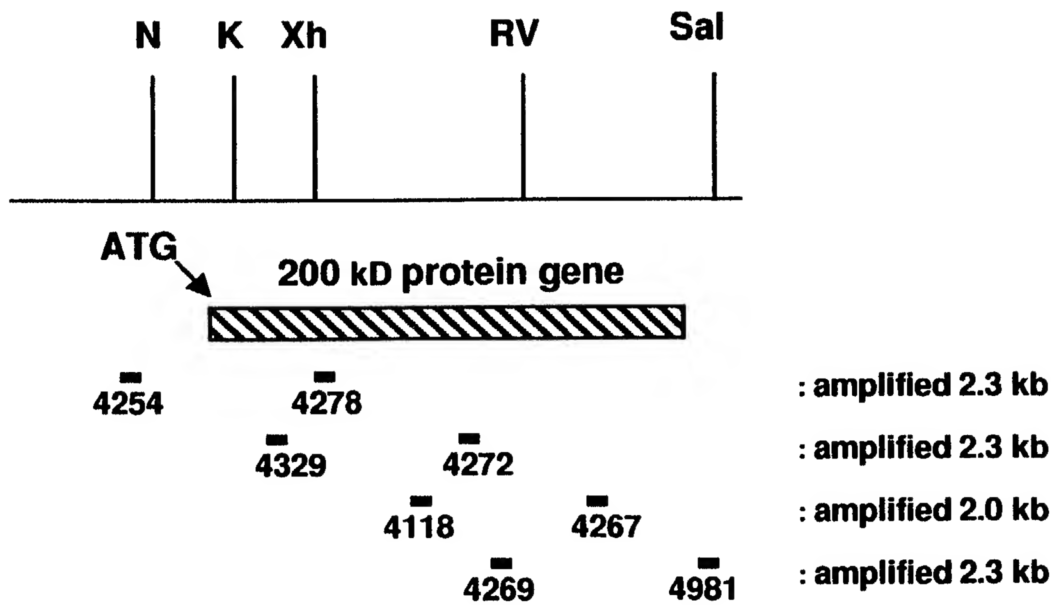


Figure 17

M. catarrhalis **strain 4223 200 kDa**

```
CCATGGATATGGGCAGGTGTGCTCGCCTGCCGTATGATGGCGATGACACCCCATTTGCCC
        10        20        30        40        50        60

CATATCTGTACGATTTGACATGTGATATGATTTAACATGTGACATGATTTAACATTGTTT
        70        80        90       100       110       120

AATACTGTTGCCATCATTACCATAATTTAGTAACGCATTTAGTAACGCATTTGTAAAAAT
       130       140       150       160       170       180

CATTGCGCCCCTTTATGTGTATCATATGAATAGAATATTATGATTGTATCTGATTATTGT
       190       200       210       220       230       240

ATCAGAATGGTGATGCTATATGATGATGCCTACGAGTTGATTTGGGTTAATCACTCTATG
       250       260       270       280       290       300

ATTTGATATATTTTGAAACTAATCTATTGACTTAAATCACCATATGGTTATAATTTAGCA
       310       320       330       340       350       360

TAATGGTAGGCTTTTTGTAAAAATCACATCGCAATATTGTTCTACTGTTACTACCATGCT
       370       380       390       400       410       420

TGAATGACGATCCCAATCACCAGATTCATTCAAGTGATGTGTTTGTATACGCACCATTTA
       430       440       450       460       470       480

CCCTAATTATTTCAATCAAATGCCTATGTCAGCATGTATCATTTTTTTAAGGTAAACCAC
       490       500       510       520       530       540

 MET ASN HIS ILE TYR LYS VAL ILE PHE ASN LYS ALA12 THR GLY THR PHE MET ALA VAL19 ALA
CATGAATCACATCTATAAAGTCATCTTTAACAAAGCCACAGGCACATTTATGGCAGTGGC
       550       560       570       580       590       600

GLU TYR ALA LYS SER HIS SER THR GLY GLY GLY SER CYS ALA THR GLY GLN VAL GLY39 SER
AGAGTACGCCAAATCCCACAGCACGGGGGGGGGTAGCTGTGCTACAGGGCAAGTTGGCAG
       610       620       630       640       650       660

VAL CYS THR LEU SER PHE ALA ARG ILE ALA ALA LEU ALA VAL LEU VAL56 ILE GLY ALA THR
TGTATGCACTCTGAGCTTTGCCCGTATTGCCGCGCTCGCTGTCCTCGTGATCGGTGCAAC
       670       680       690       700       710       720
```

FIGURE 18

3' Half Constructs of 200 kD Protein Gene

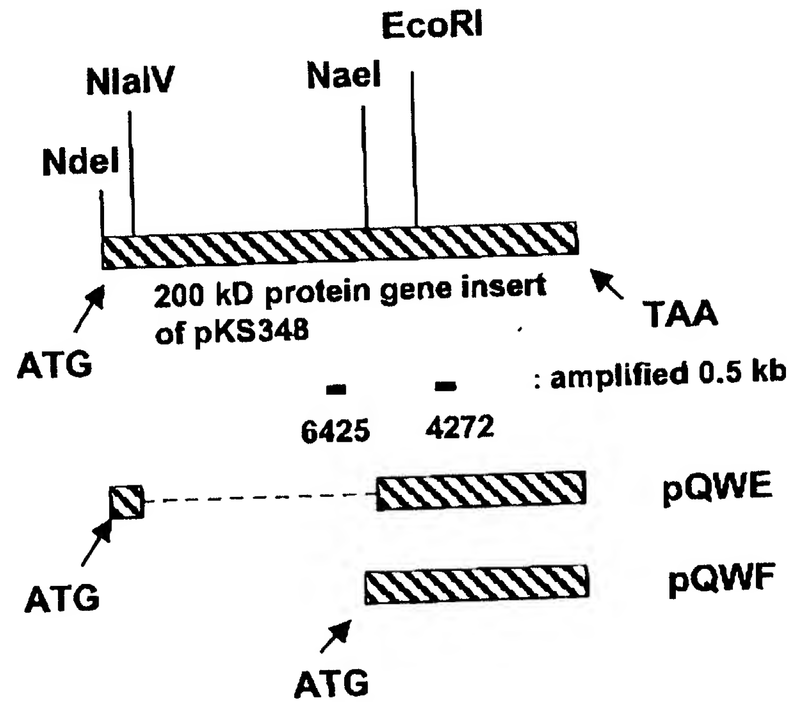


Figure 19 Construction of pQWE

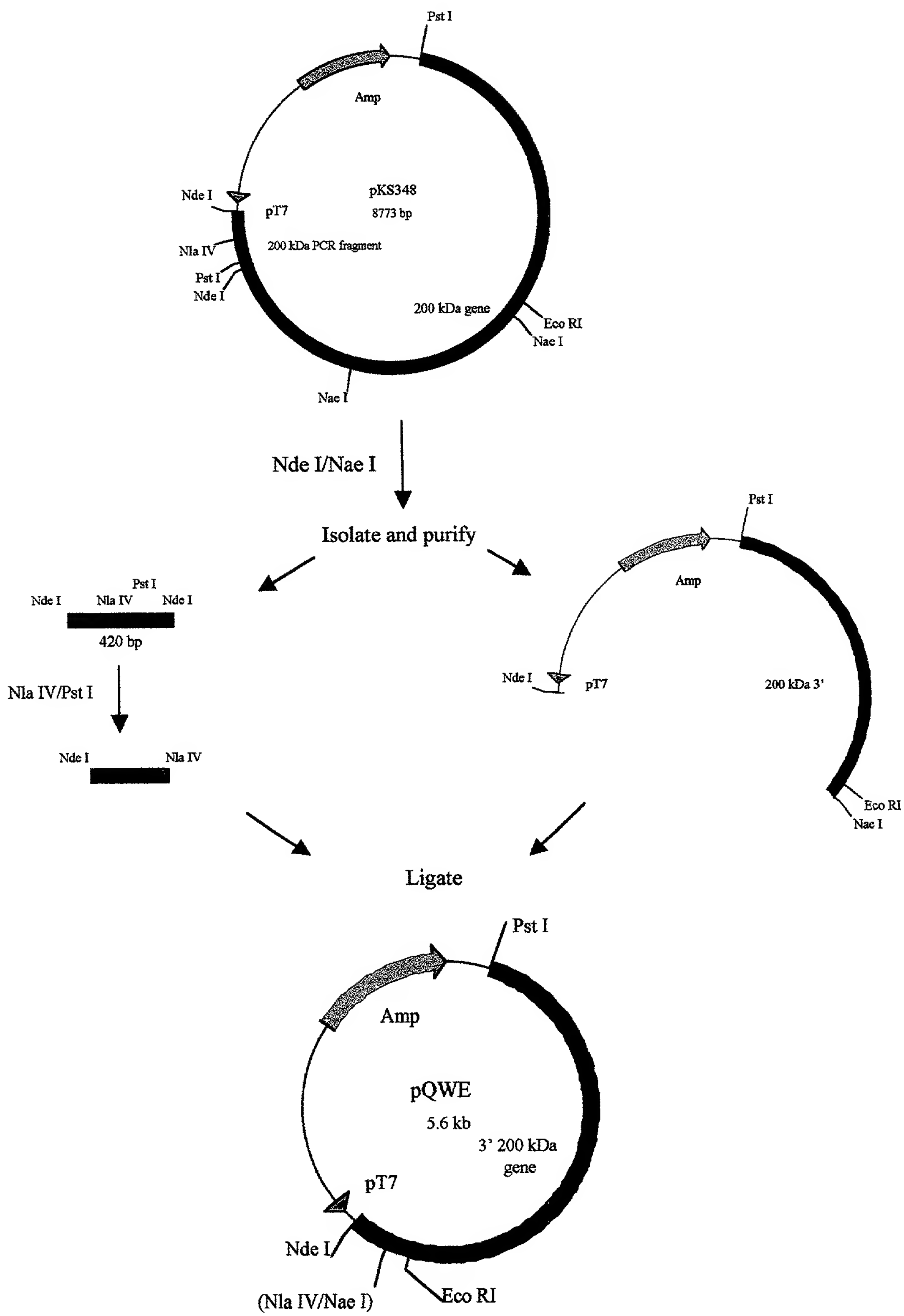
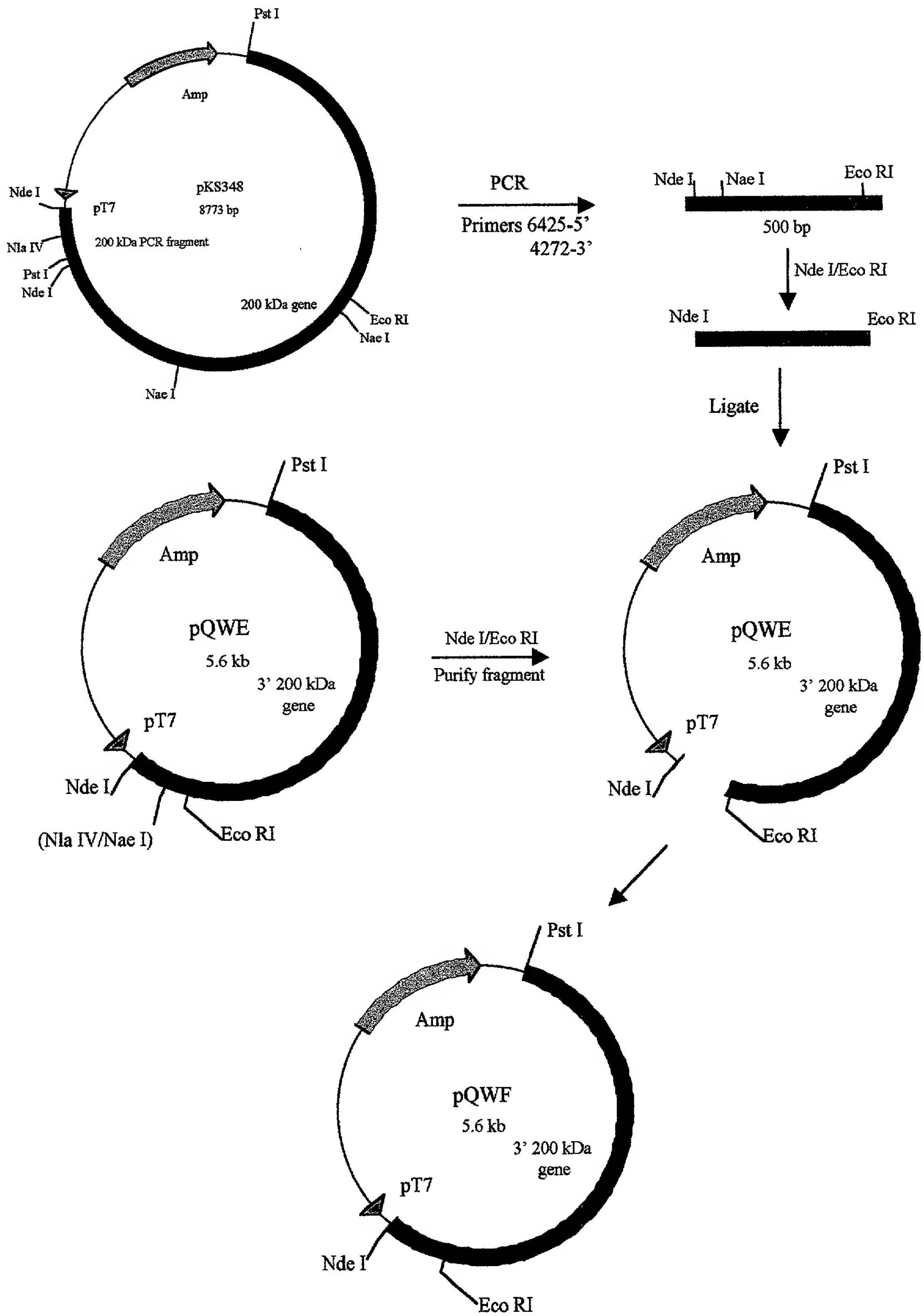


Figure 20 Construction of pQWF



Express Mail Label No.

Docket No.
1038-921 MIS:jb

Declaration and Power of Attorney For Patent Application

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF MORAXELLA

the specification of which

(check one)

☒ is attached hereto.

☐ was filed on ______________________ as United States Application No. or PCT International Application Number ______________________ and was amended on ______________________ (if applicable)

09351516.072599

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

			Priority Not Claimed
(Number)	(Country)	(Day/Month/Year Filed)	☐
(Number)	(Country)	(Day/Month/Year Filed)	☐
(Number)	(Country)	(Day/Month/Year Filed)	☐

I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

(Application Serial No.)	(Filing Date)
(Application Serial No.)	(Filing Date)
(Application Serial No.)	(Filing Date)

I hereby claim the benefit under 35 U. S. C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, C. F. R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)
(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)
(Application Serial No.)	(Filing Date)	(Status) (patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. *(list name and registration number)*

Michael I. Stewart (24,973)

Send Correspondence to: **Sim & McBurney**
6th Floor, 330 University Avenue
Toronto, Ontario
M5G 1R7, Canada.

Direct Telephone Calls to: *(name and telephone number)*
Michael I. Stewart (416) 595-1155

Full name of sole or first inventor **Sheena M. Loosmore**	
Sole or first inventor's signature	Date
Residence **Aurora, Ontario, Canada.**	
Citizenship **Canadian**	
Post Office Address **70 Crawford Rose Drive, Aurora, Ontario, Canada, L4G 4R4.**	

Full name of second inventor, if any **Ken Sasaki**	
Second inventor's signature	Date
Residence **Willowdale, Ontario, Canada.**	
Citizenship **Japanese**	
Post Office Address **Apt. 512, 1131 Steeles Avenue West, Willowdale, Ontario, Canada, M2R 3W8.**	

Full name of third inventor, if any	
Yan Ping Yang	
Third inventor's signature	Date
Residence **Willowdale, Ontario, Canada.**	
Citizenship **Canadian**	
Post Office Address **Apt. 709, 120 Torresdale Avenue, Willowdale, Ontario, Canada, M2R 3N7.**	

Full name of fourth inventor, if any	
Michel H. Klein	
Fourth inventor's signature	Date
Residence **Willowdale, Ontario, Canada.**	
Citizenship **Canadian**	
Post Office Address **16 Munro Boulevard, Willowdale, Ontario, Canada, M2P 1B9.**	

Full name of fifth inventor, if any	
Fifth inventor's signature	Date
Residence	
Citizenship	
Post Office Address	

Full name of sixth inventor, if any	
Sixth inventor's signature	Date
Residence	
Citizenship	
Post Office Address	